

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:15 ; Search time 18 Seconds

(without alignments)  
1617.068 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATRLSEEDNEDYSLN.....VLDSALRIKQSHRIDTG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	239.5	8.4	369	1	DNAJ_THEMA
2	223	7.8	326	1	DJB6_HUMAN
3	222.5	7.8	376	1	DNAJ2_AQJAE
4	211	7.4	220	1	DJB7_MOUSE
5	209.5	7.3	242	1	DJB6_MOUSE
6	204	7.1	242	1	DJB3_MOUSE
7	196	6.9	280	1	DNAJ_THERM
8	195	6.8	392	1	DNAJ_CHLPP
9	194	6.8	340	1	DJB1_HUMAN
10	193	6.8	348	1	DJB5_MOUSE
11	192.5	6.7	371	1	DNAJ_METIS
12	192.5	6.7	418	1	DNAJ2_ALIPO
13	192	6.7	337	1	DJB4_HUMAN
14	192	6.7	348	1	DJB5_HUMAN
15	191	6.7	334	1	DNAJ1_DROME
16	190	6.7	389	1	DNAJ_MYCSE
17	189	6.6	383	1	DNAJ_LACSK
18	189	6.6	390	1	DNAJ_MYCPN
19	188.5	6.6	307	1	DNAJ_DEIPR
20	188.5	6.6	397	1	DJB4_MOUSE
21	188	6.6	340	1	DJB1_MOUSE
22	187.5	6.6	373	1	DNAJ_NEIWA
23	187.5	6.6	397	1	DJB4_HUMAN
24	187	6.6	351	1	DJB2_HUMAN
25	186	6.5	337	1	DJB4_MOUSE
26	185.5	6.5	910	1	DNAJ_MYCPN
27	185	6.5	377	1	DNAJ_PSEAE
28	184.5	6.5	369	1	DNAJ_HELPY
29	184.5	6.5	369	1	DNAJ_NITEU
30	183.5	6.4	419	1	DNAJ_RHATH
31	182.5	6.4	234	1	DNAJ_RHATH
32	182.5	6.4	379	1	DNAJ_LEGPN
33	182	6.4	373	1	DNAJ_CAMOE

34	181	6.3	383	1	DNAJ_PORGI	Q9XCA6 porphyromon
35	181	6.3	437	1	DNAJ_ARRNU	P43644 atirplex nu
36	180.5	6.3	332	1	DNAJ_STYX3	P50027 synechocyst
37	180.5	6.3	392	1	DNAJ_CHLMU	Q9PK53 chlamydia m
38	180.5	6.3	392	1	DNAJ_CHLTR	O84345 chlamydia t
39	180	6.3	377	1	DNAJ_BRUME	O8Y677 bruceella me
40	180	6.3	377	1	DNAJ_BRUSU	O8FX11 bruceella su
41	180	6.3	385	1	DNAJ_CAUCR	P22305 caulobacter
42	180	6.3	387	1	DNAJ_METTS	Q9UXR9 methanocarc
43	179.5	6.3	375	1	DNAJ_BRUVU	O05980 bruceella ov
44	179.5	6.3	385	1	DNAJ_VIBHA	O87385 vibrio harv
45	179.5	6.3	392	1	NOLC_RHTRF	P26508 rhizobium f

## ALIGNMENTS

RESULT 1	ID	DNAJ_THEMA	STANDARD;	PRT;	369 AA.
AC	Q9WZV3;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-MAR-2004 (Rel. 43, Last annotation update)				
DE	Chaperone protein dnaJ.				
GN	DNAJ OR TM0849.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109 / ATCC 43589;				
RX	MEDLINE=99287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Haft D.H., Hickey B.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Uetreback T.R., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Stewart A.M., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RL	genome sequence of Thermotoga maritima."				
CC	Nature 399:323-329(1999).				
CC	-1- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE,				
CC	the ATPase activity of dnaJ (By similarity).				
CC	-1- COFACTOR: Binds 2 zinc ions per monomer (By similarity).				
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-1- SIMILARITY: Belongs to the dnaJ family.				
CC	-1- SIMILARITY: Contains 1 CR domain.				
CC	-1- SIMILARITY: Contains 1 CR domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; AE001751; AAD35931.1; --				
DR	PIR; B72327; B72327.				
DR	HSP; P08622; 1XBL.				
DR	TIGR; TM0849; --				
DR	InterPro; IPR002939; DnaJ_C.				
DR	InterPro; IPR001305; DnaJ_CXXCXGXG.				
DR	InterPro; IPR001623; DnaJ_N.				
DR	InterPro; IPR008971; HSP40_DnaJ_pep.				
DR	InterPro; IPR003095; Hsp_DnaJ.				
DR	Pfam; PF00226; DnaJ_1.				
DR	Pfam; PF01556; DnaJ_C_1.				
DR	Pfam; PF00684; DnaJ_CXXCXGXG_1.				
DR	PRINTS; PR00625; DNAJPROTEIN.				
DR	SMART; SM00271; DnaJ_1.				
DR	PROSITE; PS00636; DnaJ_1; 1.				



```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014888; BAA32209.1; -
DR EMBL; AF075601; AAD3194.1; -
DR EMBL; AF080569; AAD16010.1; ALT_FRAME.
DR EMBL; AB015798; BAA88769.1; -
DR EMBL; AB015799; BAA88770.1; -
DR EMBL; AF060703; AAP1257.1; -
DR EMBL; AL136707; CAB6642.1; -
DR EMBL; BC000177; AAH00177.1; -
DR EMBL; BC002446; AAH02446.1; -
DR HSSP; P25685; 1HDJ.
DR Genew; HGNC:14888; DNAB6.
DR GO; GO:0003773; P:heat shock protein activity; NAS.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chapterone; Alternative splicing.
DR DOMAIN; 1 69 J-DOMAIN.
DR DOMAIN; 83 172 GLY/PHE-RICH.
DR VASPLIC 232 241 VADDDALAE -> KEQLRLDNDK (in isoform B).
DR VASPLIC 242 326 Missing (in isoform B).
DR VASPLIC 242 326 /FtId=VSP_001289.
DR VASPLIC 242 326 /FtId=VSP_001290.
SQ SEQUENCE 326 AA; 36087 MW; ECF1628BF7A524F3 CRC64;

Query Match 7.8%; Score 223; DB 1; Length 326;
Best Local Similarity 32.1%; Pred. No. 1.1e-08;
Matches 67; Conservative 22; Mismatches 50; Indels 70; Gaps 7;

QY 14 DYSLVLRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFLVHQAAYVLSDP 73
DB 3 DYSLVLRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFLVHQAAYVLSDP 73
QY 74 TRAYDYVYGRKGLMEG-----WEVERRRTPAIREFERLQREERERLQORTN 124
DB 61 KRDYDYGRKGLMEG-----WEVERRRTPAIREFERLQREERERLQORTN 124
QY 125 PKGTISGVNATDLDPRDEYEDVSG-----SSPQIREINKMI 164
DB 104 -----FGGRPFSPDFDFDFDFDFDFDFDFDFDFDFDFDFDFDFDFDFDF 148
QY 165 SOSIAPLTATDTAILG-SLSTONGNGG 192
DB 149 --SFGSGFSPDFDTGFTSGSL-----GHGG 171

RESULT 3
DNJ2 AQCAE STANDARD; PRT; 376 AA.
AC 066921;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chapterone protein dnaJ-2.
GN DNAB2 OR AQ_703.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
CX NCBI_TextID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=9819666; PubMed=9537320;
Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
```

```
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL aeolicus.";
CC Nature 392:353-358(1998).
CC -1- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE,
CC the ATPase activity of dnaK (By similarity).
CC -1- COFACTOR: Binds 2 zinc ions per monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the dnaJ family.
CC -1- SIMILARITY: Contains 1 J domain.
CC -1- SIMILARITY: Contains 1 CR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000703; AAC06881.1; -
DR PIR; E70361; E70361.
DR HSSP; P08622; 1XBL.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXXCXGXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C; 1.
DR Pfam; PF00684; DnaJ_CXXCXGXG; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXXCXGXG; FALSE NEG.
DR Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KW Complete proteome.
DR Complete proteome.
DR DOMAIN; 8 72 J-DOMAIN.
DR REPEAT 156 163 CXXCXGXG MOTIF.
DR REPEAT 172 179 CXXCXGXG MOTIF.
DR REPEAT 194 201 CXXCXGXG MOTIF.
DR REPEAT 207 214 CXXCXGXG MOTIF.
DR METAL 156 156 ZINC 1 (BY SIMILARITY).
DR METAL 159 159 ZINC 1 (BY SIMILARITY).
DR METAL 172 172 ZINC 2 (BY SIMILARITY).
DR METAL 175 175 ZINC 2 (BY SIMILARITY).
DR METAL 194 194 ZINC 2 (BY SIMILARITY).
DR METAL 197 197 ZINC 2 (BY SIMILARITY).
DR METAL 207 207 ZINC 1 (BY SIMILARITY).
DR METAL 210 210 ZINC 1 (BY SIMILARITY).
SQ SEQUENCE 376 AA; 41963 MW; 9C8FBFB29A8A016D CRC64;

Query Match 7.8%; Score 222.5; DB 1; Length 376;
Best Local Similarity 26.4%; Pred. No. 1.5e-08;
Matches 102; Conservative 44; Mismatches 120; Indels 121; Gaps 19;

QY 13 EDYVSLNVRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFLVHQAAYVLSDP 72
DB 7 KDYVEILGVNRNNSQREIKAYRLVKNRDPDCKP-----CEKPKKEINAEYVLSDP 62
QY 73 QTRAYDYVYGRKGLMEG-----WEVERRRTP--AIREBF-----ERLQREERERLQORT 123
DB 63 EKRLDYDMYGAHAFEGAGAQORVETTEIPIREIILRFDFDYGSIIFERATGRRARRRR 122
QY 124 NPKG-TISVGVNATDLDPRDEYEDVSGSPQIREINKMISQSIAPLTATDTAILSG 162
DB 123 SVKGEIDVPEVET-----LEBAFK---GITVP-IEVER-----EVPQAC----- 159
QY 183 SLSTONGNGGGINFALRLRVTSAGKWELEFGAGDLOGPLFGKLPRLNLPKCFVTNNCA 242
DB 160 -----GGTGYDEBSKSRCTPTCGGRGE-----TVQGNMF--FQVRQTCPTC----- 197
```

```

OY 243 LQFSSRGIRPGLTTTARNDKNT-VGYLOWRMGOSAMNTSVPRDTKSHFTVALQGI 301
DB 198 -----GSEGVYENHACTGRGY-----GLVKET-----IKVKI 226
OY 302 PHSFALISYGHKQDDQGTIVKSLKAFRC-----TYVEGAERKISR----- 345
DB 227 PPG-----VRDSKLVEGKHAGRGYGGPPGLYITIVKPKHIFERKDDLYVDV 277
OY 346 -----HSLGAASVGPQGVSLKVKL 367
DB 278 NITYPEAVLCTEVEVPTLDEKXKVKI 304

RESULT 4
DJB7 MOUSE STANDARD; PRT; 220 AA.
AC 09OY18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 7 (mdj5).
GN DnaJb7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
RX MEDLINE=21023480; PubMed=11147971;
RA Ohtsuka K., Hata M.;
RT "Mammalian HSP40/DnaJ homologs: cloning of novel cDNAs and a proposal
RL for their classification and nomenclature.";
CC Cell Stress Chaperones 5:98-112(2000).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB028855; BAA8303.1; -.
CC DR HSP; P25685; 1HDJ.
CC DR MGD; MGI:1914012; DnaJb7.
CC DR InterPro; IPR001623; DnaJ_N.
CC DR Pfam; PF00226; DnaJ; 1.
CC DR PROSITE; PS00636; DnaJ_1; 1.
CC DR PROSITE; PS50076; DnaJ_2; 1.
CC KM Chapterone.
CC FT DOMAIN 3 69 J-DOMAIN.
CC SQ SEQUENCE 220 AA; 25530 MW; E741D9232713F7E CRC64;

Query Match 7.4%; Score 211; DB 1; Length 220;
Best Local Similarity 33.3%; Pred. No. 4.7e-08;
Matches 59; Conservative 25; Mismatches 49; Indels 44; Gaps 6;

OY 14 DYISLANVREASSEELKAAVRLCMYHPDKRDEPLKSOARLRLVHQAVEYSDPO 73
DB 3 DYEEVAGVQRYAPEDIKRAYRVALKWHPDK--NPNENKEARKEKVEAAEYVLSNVE 60
OY 74 TRAIYDIYGRKGEMEGMEVEVRRRPAEIRFEERLQREERERRLQQRTPNGTISVGV 133
DB 61 KRDIYDKYKGGSDGKRG-----ASHLDREKER----- 88
OY 134 DATDLFDREYDEEYDVSGSSFPQIEINKMI-SQSIAPPLATATDAILSGSLSTONG 189
DB 89 ---FTFRKADVDVEKEIFGERD-----FSFLLFEDSLGGLNSRTP--SGSRERRAG 136

```

```

DJB6 MOUSE STANDARD; PRT; 242 AA.
AC 054946; O9OY18;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DnaJ homolog subfamily B member 6 (Heat shock protein J2) (HSJ-2)
DE (MRJ) (mdj4).
GN DnaJb6 OR HSJ2 OR MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Trophoblast;
RA Hunter P.J., Swanson B.U., Haendel M., Lyons G.E., Cross J.C.;
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=21023480; PubMed=11147971;
RA Ohtsuka K., Hata M.;
RT "Mammalian HSP40/DnaJ homologs: cloning of novel cDNAs and a proposal
RL for their classification and nomenclature.";
CC Cell Stress Chaperones 5:98-112(2000).
CC [3]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=22388257; PubMed=12477932;
CC RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
CC RA Altshuler R.D., Collins F.S., Wagner L., Scheffen C.M., Schuler G.D.,
CC RA Ransone S.J., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
CC RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
CC RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
CC RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
CC RA Brownstein M.T., Ueda T.B., Toshnyuk S., Carninci P., Prange C.,
CC RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
CC RA Bock S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
CC RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
CC RA Villalon D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
CC RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
CC RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
CC RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
CC RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
CC RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
CC RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
CC RT "Generation and initial analysis of more than 15,000 full-length
CC human and mouse cDNA sequences.";
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF035962; AAC16759.1; -.
CC DR EMBL; AB028854; BAA8302.1; -.
CC DR EMBL; BC003702; AAH03702.1; -.
CC DR HSP; P25685; 1HDJ.
CC DR MGD; MGI:1344381; DnaJb6.
CC DR InterPro; IPR001623; DnaJ_N.
CC DR Pfam; PF00226; DnaJ; 1.
CC DR SMART; SM00271; DnaJ; 1.
CC DR PROSITE; PS00636; DnaJ_1; 1.
CC DR PROSITE; PS50076; DnaJ_2; 1.
CC KM Chapterone.
CC FT DOMAIN 1 69 J-DOMAIN.
CC FT DOMAIN 83 170 GUY/PHE-RICH.
CC FT CONFLICT 139 140 GS -> AP (IN REF. 1).

```



```

FT CONFLICT 218 218 E -> V (IN REF. 3).
FT CONFLICT 227 227 S -> P (IN REF. 1).
SQ SEQUENCE 242 AA; 26978 MM; 3856053CJB0735D CRC64;

Query Match
Best Local Similarity 3.0%; Score 209.5; DB 1; Length 242;
Matches 65; Conservative 21; Mismatches 53; Indels 71; Gaps 7;

QY 14 DYSLNVRRAASSEELKAAVRLCMLYHPDKRDPDLKSAEPLFVLVQAAYEVSDPQ 73
DB 3 DYVVLGVQORASPEDIKKAYKQALKWHDPK--NPEKKEAEKRFQVAVAEVLSDAK 60
QY 74 TRAIYDIYGRKGLMEG-----WEVERRRTPAIREERERLQREERERLQORT 123
DB 61 KRDIYDKYKGGKGLNGGGGGGSIHFDSPFEFGFTFRNPDDVREF----- 104
QY 124 NPKGTISGVDAITDLPFRYDEYDVSG-----SSFPQIEINKMH 163
DB 105 -----FGGRDPSFDFPFDFPDFFGNRRGPRGNRSGAGSPFSTPSGPF----- 149
QY 164 ISOSIEAPLTATDTAITS--GSLSTQNGNG 192
DB 150 ---SFGSGFPAFDGTPTPFGSL---GHGG 172

RESULT 6
DJB3_MOUSE STANDARD; PRT; 242 AA.
ID DJB3_MOUSE
AC 035723; Q9DAN3; Q9DAN4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DnaJ homolog subfamily B member 3 (DnaJ protein homolog 3) (Heat shock
U3 protein) (HSC-3) (MSJ-1).
GN DNABJ3 OR HSC3 OR MSJ1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98189155; PubMed=9521861;
RA Berruti G., Perigo L., Borgonovo B., Martegani E.;
RT "MSJ-1, a new member of the DnaJ family of proteins, is a male germ
cell-specific gene product.";
RL Exp. Cell Res. 239:430-441(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli J., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gaitanidis M.,
RA Guernicich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaurer P.,
RA Nordone P., Ring B., Schenwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: May operate as a co-chaperone of the male germ cell- and
haploid stage-specific Hsp70 proteins.

```

```

CC -1- TISSUE SPECIFICITY: Testis specific. Expression is confined to the
CC germ line without any contribution of the somatic components.
CC -1- DEVELOPMENTAL STAGE: Its expression occurs in the postmeiotic
CC phase of male germ cell development. First detected in 30 days old
CC mice and thereafter into adulthood. Barely detectable in 20 days
CC old mice and absent before this period.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U95607; AAC13944.1; -.
DR EMBL; AK005690; BAB24188.1; -.
DR EMBL; AK005688; BAB24186.1; -.
DR HSSP; P25685; 1HDI.
DR MGD; MGI:1306822; DnaJb3.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KM Chapterone.
FT DOMAIN 1 69 J-DOMAIN.
FT CONFLICT 71 71 V -> M (IN REF. 2; BAB24186).
FT CONFLICT 149 149 F -> I (IN REF. 2; BAB24188).
FT CONFLICT 233 233 R -> K (IN REF. 2).
SQ SEQUENCE 242 AA; 26678 MM; 5F54P48A00094D84 CRC64;

Query Match
Best Local Similarity 30.6%; Pred. No. 1.7e-07;
Matches 63; Conservative 31; Mismatches 64; Indels 48; Gaps 7;

QY 14 DYSLNVRRAASSEELKAAVRLCMLYHPDKRDPDLKSAEPLFVLVQAAYEVSDPQ 73
DB 3 DYVVLGVQORASPEDIKKAYKQALKWHDPK--NPEKKEAEKRFQVAVAEVLSDAK 60
QY 74 TRAIYDIYGRKGLMEG-----WEVERRRTPAIREERERLQREERERLQORT 108
DB 61 KREYIDRCGVG--EVGGGGAAGSPFDPAFYVSPFDPAVFRFFGCHDPSPDFRGCD 119
QY 109 RLQREERERLQORTPKGTISGVDAITDLPFRYDEYDVSGSSPPQIEINKMHISQST 168
DB 120 PLENFEDRSTRSGSRNRRGAVPSTSPTE-FPGF-----CGGF-----ASL 159
QY 169 EAPLTATDTAITS--GSLSTQNGNGGS 194
DB 160 DTGFTSFGSPNGSLSPFMSCGGA 185

RESULT 7
DnaJ_THETH STANDARD; PRT; 280 AA.
ID DnaJ_THETH
AC Q56237; P77642;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein dnaJ.
GN DnaJ.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermaleae; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=98007877; PubMed=9349721;
RA Osipluk J., Joachimiak A.;
RT "Cloning, sequencing, and expression of dnaJ-operon proteins from the

```

RT thermophilic bacterium *Thermus thermophilus*.";   
 RL Biochim. Biophys. Acta 1353:253-265(1997).   
 RN [2]   
 RP SEQUENCE FROM N.A.   
 RC STRAIN=HB8 / ATCC 27634;   
 RA Seidel R.;   
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.   
 RN [3]   
 RP SEQUENCE FROM N.A.   
 RC STRAIN=HB8 / ATCC 27634;   
 RX MEDLINE=9914855; PubMed=10092456;   
 RA Klostermeier D., Seidel R., Reinsteijn J.;   
 RL "The functional cycle and regulation of the *Thermus thermophilus* DnaK chaperone system.";   
 RT J. Mol. Biol. 287:511-525(1999).   
 RL [4]   
 CC -1- FUNCTION: Does not influence ATP binding or hydrolysis nor ADP release. Exerts influence on the interaction of DnaK with substrates; in the presence of Dafa, DnaJ inhibits substrate binding, and substrate already bound to DnaK is displaced by DnaJ and Dafa.   
 CC -1- SUBUNIT: Forms a heteronamer with DnaJ and Dafa in the resting state. Three copies of each protein are present in the complex.   
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).   
 CC -1- SIMILARITY: Belongs to the dnaJ family.   
 CC -1- SIMILARITY: Contains 1 J domain.   
 CC -----   
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).   
 CC -----   
 CC EMBL; L57504; AAB04678.1; -   
 DR EMBL; Y07826; CAA69161.1; -   
 DR EMBL; D84222; BAA1282.1; -   
 DR EMBL; AB012390; BAA81743.1; -   
 DR EMBL; AB032368; BAA96087.1; -   
 DR HSSP; P25685; 1HDJ.   
 DR InterPro; IPR002839; DnaJ\_C.   
 DR InterPro; IPR001623; DnaJ\_N.   
 DR InterPro; IPR003095; Hsp\_DnaJ.   
 DR Pfam; PF00226; DnaJ\_1.   
 DR Pfam; PF01556; DnaJ\_C\_1.   
 DR PRINTS; PR00625; DNAJPROTEIN.   
 DR SMART; SM00271; DnaJ\_1.   
 DR PROSITE; PS00636; DNAJ\_1; 1.   
 DR PROSITE; PS50076; DNAJ\_2; 1.   
 KW Chaperone; DNA replication; Heat shock.   
 FT DOMAIN 4   
 FT DOMAIN 72   
 FT DOMAIN 75   
 FT DOMAIN 80   
 FT POLY-PRO.   
 FT CONFLICT 129 129 E -> K (IN REF. 1).   
 FT SEQUENCE 280 AA; 30978 MW; AE32D595359388D0 CRC64;

```

Query Match      6.9% Score 196; DB 1; Length 280;
Best Local Similarity 36.2%; Pred. No.7.7e-07;
Matches 46; Conservative 20; Mismatches 31; Indels 30; Gaps 4;

QY    13 EDVYSLLNVRREASSEELKAAAYRRLCMLYHDKHRDPELKSQAERLFNLVHQAEVYISDP   72
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db     5 KDYYAILGVPRNNATQGEIKRAYRKRLARQYHDVNKSPF---AEEKFKEINBAVAVISDP   60
       :|||:|||::-----GLEMEGHEVVVERRRTPMIEEFEBRL-----ORE 113

```

[illegible]



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X62421; CA444287.1; -  
DR EMBL; D49547; BAA08495.1; -  
DR EMBL; D85429; BAA12819.1; -  
DR EMBL; BC002352; AAH02352.1; -  
DR EMBL; BC019827; AAH19827.1; -  
DR PIR; JN0912; JN0912.  
DR PIR; S20062; S20062.  
DR PDB; 1HDJ; 08-NOV-96.  
DR Genew; HGNC:5270; DNABJ1.  
DR MIM; 604572; -  
DR GO; 0003773; F:heat shock protein activity; TMS.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR008971; Hsp40\_DnaJ\_pep.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF01556; DnaJ\_C.1.  
DR PRINTS; PR00625; DnaJPROTEIN.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PSS0076; DnaJ\_2; 1.  
KM Heat shock; Chaperone; 3D-structure.  
FT DOMAIN 1 70  
FT CONFLICT 11 11 L -> Q (IN REF. 1).  
FT CONFLICT 13 28 RGASDEIKRAYRROA -> ALGRGDQAGLPPPG (IN  
REF. 1).  
FT CONFLICT 68 68 G -> L (IN REF. 1).  
FT CONFLICT 81 136 SGGANCTSPSYTHFDPMFAFFGRRNPPTFGQNG  
EEGMDIDPPSGFPM -> TAEPMPVPLGATHMETLMPCL  
LSSVAEIPLPPLPGSGTGKAWTLMTHTSLIIM (IN  
REF. 1).  
FT CONFLICT 150 150 R -> C (IN REF. 1).  
FT CONFLICT 183 183 M -> T (IN REF. 1).  
FT CONFLICT 320 320 V -> A (IN REF. 1).  
FT HELIX 6 9  
FT TURN 10 10  
FT TURN 13 14  
FT HELIX 17 29  
FT TURN 30 31  
FT TURN 33 35  
FT TURN 39 40  
FT HELIX 41 54  
FT TURN 55 56  
FT HELIX 58 66  
FT TURN 67 67  
FT HELIX 69 71  
SQ SEQUENCE 340 AA; 38044 MW; 17545098B0C196DF CRC64;

Query Match 6.8%; Score 194; DB 1; Length 340;  
Best Local Similarity 48.7%; Pred. No. 1.4e-06;  
Matches 38; Conservative 14; Mismatches 22; Indels 4; Gaps 1;

QY 13 EDVYSLNVRKASSEELKAYRRLCLYPPDKGRPELKSQAEKRLFNLVHQAIVYLSLP 72  
DB 3 KDYYQTGLARGASDEIKRAYRROALRYHPDKNEP---GAEEKFEIAEAYVYLSLP 58  
QY 73 QTRAIYDIYGRKGLIEMEG 90  
DB 59 RKREIPRYGEELKSGG 76

RESULT 10  
DUBS\_MOUSE STANDARD; PRT; 348 AA.  
AC 089114;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE DnaJ protein subfamily B member 5 (Heat shock protein Hsp40-3) (Heat  
shock protein cognate 40) (Hsc40).  
GN DNABJ5 OR HSC40.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20035738; PubMed=10570961;  
RT Chen M.-S., Roti J.R., Laszlo A.;  
RT "Hsc40, a new member of the hsp40 family, exhibits similar expression  
RT profile to that of hsc70 in mammalian cells.";  
RL Gene 238:333-341(1999).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loughell N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- INDUCTION: Expressed under normal conditions, its expression can  
CC further be increased after various stress treatments.  
CC -1- SIMILARITY: Contains 1 J domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF092535; AAC64141.1; -  
DR EMBL; AF088983; AAC35861.1; -  
DR EMBL; AF321322; AAG53972.1; -  
DR EMBL; BC057087; AAH57087.1; -  
DR HSSP; P25685; 1HDJ.  
DR MGD; MG1:1930018; DnaJb5.  
DR InterPro; IPR002939; DnaJb5.  
DR InterPro; IPR001623; DnaJ\_C.  
DR InterPro; IPR008971; Hsp40\_DnaJ\_pep.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C.1.  
DR PRINTS; PR00625; DnaJPROTEIN.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PSS0076; DnaJ\_2; 1.  
KM Chaperone.  
FT DOMAIN 4 68 J-DOMAIN.  
SQ SEQUENCE 348 AA; 39119 MW; 833D2F5547321687 CRC64;

Query Match 6.8%; Score 193; DB 1; Length 348;  
Best Local Similarity 43.6%; Pred. No. 1.7e-06;  
Matches 34; Conservative 20; Mismatches 20; Indels 4; Gaps 1;

[illegible]

```

SQ  SEQUENCE 371 AA; 41052 MW; 86CAAF5EB87CF95C CRC64;
Query Match 6.7%; Score 192.5; DB 1; Length 371;
Best Local Similarity 50.7%; Pred. No. 2e-06;
Matches 38; Conservative 12; Mismatches 22; Indels 3; Gaps 1;

QY 13 EDVYSILNVRREASSEELKAAVRLICMLVHPDKRDELKSAQERLFLNVLHQAYEVLSDP 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4 KDYYEVLGVNRDSDSEESIKYSKRLAMKHPDRNP--NPKAESFEAKAEAYEVLSD 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 73 QTRAIVDIYGRKGLE 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 QKRAAYDQYGHAGVD 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
DNU2_ALIPO STANDARD; PRT; 418 AA.
AC P42824.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DnaJ protein homolog 2.
GN LDJ2.
OS Allium porrum (leek).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
OC Allium.
OX NCBI_TaxID=4681;
RN [1]
RP SEQUENCE FROM N.A.
RA Beesoule J.J., Testet E., Caesagne C.;
RT "Cloning of a new isoform of a DnaJ protein from Allium porrum
RL epidermal cells.";
RL Plant Physiol. Biochem. 33:723-727(1994).
CC -1- FUNCTION: Have a continuous role in plant development probably in
CC the structural organization of compartments (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77632; CAA54720.1; -.
DR PIR; S42031; S42031.
DR HSSP; P25685; IHDJ.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXXCXGXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pep.
DR InterPro; IPR003095; Hep_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C_1.
DR Pfam; PF00684; DnaJ_CXXCXGXG_1.
DR PRINTS; PR00625; DNADJPROTEIN.
DR SMART; SMO0271; DnaJ_1.
DR SMART; PS00636; DnaJ_1; 1.
DR PROSITE; PS00637; DnaJ_CXXCXGXG_1.
DR PROSITE; PS50076; DnaJ_2; 1.
KW Chaperone; Repeat; Prenylation; Lipoprotein; Multigene family.
FT DOMAIN 11
FT DOMAIN 88
FT DOMAIN 81 108 GLY-RICH.
FT REPEAT 148 155 CXXCXGXG MOTIF.
FT REPEAT 164 171 CXXCXGXG MOTIF.
FT REPEAT 191 198 CXXCXGXG MOTIF.
FT REPEAT 207 214 CXXCXGXG MOTIF.

```

FT LIPID 415 415 S-farnesyl cysteine (By similarity).  
 SQ SEQUENCE 418 AA; 46584 MW; DCE2A4DF192329E6 CRC64;  
 Query Match 6.7%; Score 192.5; DB 1; Length 418;  
 Best Local Similarity 29.4%; Pred. No. 2,4e-06;  
 Matches 65; Conservative 30; Mismatches 69; Indels 57; Gaps 9;  
 QY 11 DNEEDYSLNVRREASSEELKAAYRRLCMLYHPDKHDPKLSQAERLFNVLHQAAYEVL 70  
 DB 10 DNTKYIVLVGSKATPEDLKAYRKALIKHPKGDPE-----KFEIGQAEVLN 62  
 QY 71 DPQTRAIYDIYKRGKLEMEGWEVVERRRTPAIREEF-----ERLQREERERLQQR 122  
 DB 63 DPEKREIYDYGEGELK-EGWGGGGVHDPFDIFQSPFGGSGGSSRGRRQREGEDV 121  
 QY 123 TNPKGTISVGVADTDLFDRLDEDEVDVSGSSPQIEINKNHIGOSIAPLTADTALISG 182  
 DB 122 VHP-----LKVSELDLY-----NGTS-----KKLISLRN-----VLCT 149  
 QY 183 SLSTONGGGGGINFALRVTSAGWGE-----LEFGAGDLO 219  
 DB 150 KCKGKSGSKSGAMNCA-----SCGSGMKVSIROLGFGMIQ 185  
 RESULT 13  
 DJB4 HUMAN STANDARD; PRT; 337 AA.  
 AC Q9UDF4; Q13431; Created  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DnaJ homolog subfamily B member 4 (Heat shock 40 kDa protein 1 homolog) (Heat shock protein 40 homolog) (Hsp40 homolog).  
 GN DNABJ4 OR HJL1 OR DNABJW.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98207666; PubMed=9546042;  
 RA Hoe K.L., Won M., Chung K.S., Jang Y.J., Lee S.B., Kim D.U., Lee J.W., Yun J.H., Yoo H.S.;  
 RT "Isolation of a new member of DnaJ-like heat shock protein 40 (Hsp40) from human liver".  
 RL Biochim. Biophys. Acta 1383:4-8(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliaty S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP PRELIMINARY SEQUENCE FROM N.A.

RC TISSUE=Liver;  
 RA Won M., Moon K.M., Lee C.E., Yoo H.S.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DDAJ databases.  
 CC -1- INDUCTION: By heat shock.  
 CC -1- SIMILARITY: Contains 1 J domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; UA0992; AAC14483.2; -  
 DR EMBL; BC034721; AAH34721.1; -  
 DR EMBL; U41290; AAB07346.1; ALT\_FRAME.  
 DR HSSP; P25685; 1HDJ.  
 DR Genew; HGNC:14886; DNABJ4.  
 DR GO; GO:0003773; P:heat shock protein activity; TAS.  
 DR GO; GO:000408; P:response to heat; TAS.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR008971; Hsp40\_DnaJ\_pdp.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR Pfam; PF01566; DnaJ\_C; 1.  
 DR PRINTS; PR00625; DNABPROTEIN.  
 DR PROSITE; PS00636; DNAB\_1; 1.  
 DR PROSITE; PS0076; DNAB\_2; 1.  
 KW Chaperone; Heat shock.  
 FT DOMAIN 1 70  
 SQ SEQUENCE 337 AA; 37806 MW; C7A9C613F73BCDAC CRC64;  
 Query Match 6.7%; Score 192; DB 1; Length 337;  
 Best Local Similarity 48.0%; Pred. No. 1.9e-06;  
 Matches 36; Conservative 16; Mismatches 19; Indels 4; Gaps 1;  
 QY 13 EDYSSLNVRREASSEELKAAYRRLCMLYHPDKHDPKLSQAERLFNVLHQAAYEVLSP 72  
 DB 3 KDYYCTLIGEKGSDDIDKAYRKALIKHPKNSP-----QAEKFEVAEAYEVLSP 58  
 QY 73 QTRAIYDIYKRGKLE 87  
 DB 59 KKEIREIDYGEELK 73  
 RESULT 14  
 DJB5 HUMAN STANDARD; PRT; 348 AA.  
 ID DJB5 HUMAN 075953; O8TDR7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DnaJ homolog subfamily B member 5 (Heat shock protein 40) (Heat shock protein cognate 40) (Hsc40) (Hsp40-2).  
 GN DNABJ5 OR HSC40.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=20035738; PubMed=10570961;  
 RA Chen M.-S., Roti J.R., Laszlo A.;  
 RT "Hsc40, a new member of the hsp40 family, exhibits similar expression profile to that of hsc70 in mammalian cells".  
 RL Gene 238:333-341(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Fu Q., Yu L., Yue P., Zhou Y., Jiang J.X., Zhao S.Y.;  
 RT "Cloning and expression of a new human cDNA homology to human heat-

```

RT shock protein 40 mRNA.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- INDUCTION: Expressed under normal conditions, its expression can
CC further be increased after various stress treatments.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF088982; AAC35860.1; -.
CC EMBL; AF087870; AAM10498.1; -.
CC HSSP; P25685; 1HDJ.
CC Genew; HGNC:14887; DNAJB5.
CC GO; GO:0003773; F:heat shock protein activity; TAS.
CC GO; GO:0006950; P:response to stress; TAS.
CC InterPro; IPR002939; DnaJ_C.
CC InterPro; IPR001623; DnaJ_N.
CC InterPro; IPR008971; HSP40_DnaJ_pep.
CC InterPro; IPR003095; Hsp_DnaJ.
CC Pfam; PF00226; DnaJ_1.
CC Pfam; PF01556; DnaJ_C_1.
CC PRINTS; PR00625; DNAJPROTEIN.
CC PROSITE; PS00636; DnaJ_1; 1.
CC PROSITE; PSS0076; DnaJ_2; 1.
CC Chaperone.
CC DOMAIN
FT DOMAIN 4 68 MKITRRR -> IEDHKAS (IN REF. 2).
FT CONFLICT 190 196 TP -> HL (IN REF. 2).
FT CONFLICT 234 235
SQ SEQUENCE 348 AA; 39133 MW; DC9FG45DE4PF8CFC CRC64;

Query Match 6.7%; Score 192; DB 1; Length 348;
Best Local Similarity 43.6%; Pred. No. 2e-06;
Matches 34; Conservative 19; Mismatches 21; Indels 4; Gaps 1;

QY 13 EDYSLINVRREASEELKAYRRLCMLYHDKRDELNSQAEFLNLYHQAYVLSDP 72
DB 3 KDYIKKIGIPSGANEDIKKAYRGMALKYHDKKEP-----NMEKFEKIEAAYDVLSDP 58
QY 73 QTRAIYIYKRGKLEMEG 90
DB 59 KKGGLYDQYGEGLKGTG 76

RESULT 15
DNU1_DROME STANDARD; PRT; 334 AA.
AC Q24133; Q9VRP0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DnaJ protein homolog 1 (DROJ1).
GN DnaJ-1 OR DROJ1 OR CG10578.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
OX
RN [1]
RP SEQUENCE FROM N.A.
RA Lee J.Y., Palter K.B.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20156006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

```

```

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Aoril J.F., Agayanti A., An H.-O., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA Borokova D., Bosman M.R., Bouck J., Brokstein P., Brotier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TSSU=Embryo;
RX MEDLINE=22426066; PubMed=12537369;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
RA Rubin G.M., Celnikier S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: By heat shock.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U34904; AAC23584.1; -.
CC EMBL; AB003565; AAF50753.1; -.
CC EMBL; AY058788; AAL14017.1; -.
CC HSSP; P25685; 1HDJ.
CC Genew; HGNC:14887; DNAJB5.
CC GO; GO:0003773; F:heat shock protein activity; TAS.
CC GO; GO:0006950; P:response to stress; TAS.
CC InterPro; IPR002939; DnaJ_C.
CC InterPro; IPR001623; DnaJ_N.
CC InterPro; IPR008971; HSP40_DnaJ_pep.
CC InterPro; IPR003095; Hsp_DnaJ.
CC Pfam; PF00226; DnaJ_1.
CC Pfam; PF01556; DnaJ_C_1.
CC PRINTS; PR00625; DNAJPROTEIN.
CC SMART; SM00271; DnaJ_1.
CC PROSITE; PS00636; DnaJ_1; 1.
CC PROSITE; PSS0076; DnaJ_2; 1.
CC Chaperone; Heat shock.
FT DOMAIN 4 68

```



FT	CONFLICT	261	261	G -> E (IN REF. 1).
SQ	SEQUENCE	334	AA; 37028 MW; 60542ABFD47A5689 CRC64;	

SQ SEQUENCE 334 AA; 37028 MW; 60542ABFD47A5689 CRC64;

Query Match	6.7%;	Score 191;	DB 1;	Length 334;
Best Local Similarity	48.0%;	Prod No 3	20-06-	

Best local similarity 50.0%; Freq. NO. 2.4E+06;  
Matches 36; Conservative 17; Mismatches 18; Gaps 1;

Best local similarity 50.0%; Freq. NO. 2.4E+06;  
Matches 36; Conservative 17; Mismatches 18; Gaps 1;

```
QY      13 EDYSLINVRKASSEEIKAAVRLCMLYHPDKRDELPKLSQAERLFNLVHQAAYEYLSDP 72
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      3 KDFYKILGLERRASDDEIKAAVRLATLKVHPDNKSP-----QAERFKELAEAYEYLSDK 58
```

Db 3 KDFYKILGLERKASDDEIKAYRKLALKTHPDKNKSP---QAEERFKEIAAEAYEVLSDK 58

QY 73 QTRAIYDIYCKRGLE 87

Db 59 KKRDI<sup>+</sup>FDNYGGEDGLK 73

Search completed: July 13, 2004, 08:41:07  
Job time : 19 secs

Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:15 ; Search time 79 Seconds  
(without alignments)  
1999.292 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATRALSEELDNDYSLN.....VLDSALRIPKSHRIDTDG 559

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: \_geneseqp1980s:\*  
2: \_geneseqp1990s:\*  
3: \_geneseqp2000s:\*  
4: \_geneseqp2001s:\*  
5: \_geneseqp2002s:\*  
6: \_geneseqp2003as:\*  
7: \_geneseqp2003bs:\*  
8: \_geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2857	100.0	559	3 AAY91945	AAY91945 Human cha
2	2857	100.0	559	5 ABG79329	ABG79329 Human foe
3	2848	99.7	559	4 AAB93115	AAB93115 Human pro
4	1723	60.3	336	5 ABG79328	ABG79328 Human foe
5	1408	49.3	356	4 AAM39772	AAM39772 Human pol
6	1408	49.3	360	4 AAM41558	AAM41558 Human pol
7	1401	49.0	360	4 ABG14697	ABG14697 Novel hum
8	1128.5	39.5	521	4 ABB63918	ABB63918 Drosophil
9	1124	39.3	344	4 ABG03199	ABG03199 Novel hum
10	750	26.3	159	4 AAU17215	AAU17215 Novel sig
11	750	26.3	159	7 ADB93923	ADB93923 Human nov
12	733	25.7	159	4 AAU17589	AAU17589 Novel sig
13	733	25.7	159	7 ADB94297	ADB94297 Human nov
14	676.5	23.7	147	4 ABG14695	ABG14695 Novel hum
15	613	21.5	119	3 AAG01815	AAG01815 Human sec
16	532	18.6	149	3 AAB43143	AAB43143 Human ORF
17	456.5	16.0	118	3 AAB07861	AAB07861 Amino aci
18	451	15.8	103	5 ABG79330	ABG79330 Human foe
19	412.5	14.4	253	4 ABG14698	ABG14698 Novel hum
20	412.5	14.4	253	4 ABB03200	ABB03200 Novel hum
21	301	10.5	64	4 ABG03197	ABG03197 Novel hum
22	235.5	8.2	328	3 AAB18261	AAB18261 Plasmodiu
23	226.5	7.9	349	3 AAG25397	AAG25397 Arabidops
24	224.5	7.9	349	3 AAG42820	AAG42820 Arabidops
25	223	7.8	242	5 ABP41542	ABP41542 Human ova

26	223	7.8	317	2 AAY74126	AAY74126 Human pro
27	223	7.8	321	5 ABP69546	ABP69546 Human pol
28	223	7.8	330	2 AAM94066	AAM94066 Human DNA
29	218	7.6	354	3 AAB18160	AAB18160 Plasmodiu
30	211	7.4	337	3 AAG16066	AAG16066 Arabidops
31	208	7.3	343	3 AAG13707	AAG13707 Arabidops
32	204	7.1	348	3 AAG30654	AAG30654 Arabidops
33	204	7.1	350	3 AAG53766	AAG53766 Arabidops
34	204	7.1	350	3 AAG18156	AAG18156 Arabidops
35	203	7.1	333	3 AAG18157	AAG18157 Arabidops
36	203	7.1	333	3 AAG53767	AAG53767 Arabidops
37	201	7.0	335	3 AAG13708	AAG13708 Arabidops
38	195	6.8	392	2 AAY34630	AAY34630 Chlamydia
39	194	6.8	340	4 AAB72675	AAB72675 Human HDJ
40	194	6.8	340	5 AAU80168	AAU80168 Human hea
41	194	6.8	377	4 ABG18525	ABG18525 Novel hum
42	193	6.8	348	7 ADB59551	ADB59551 Rat Prote
43	193	6.8	402	7 ABB73980	ABB73980 DNA clone
44	192	6.7	304	3 AAY91943	AAY91943 Human cha
45	192	6.7	336	4 AAB62196	AAB62196 Amino aci

## ALIGNMENTS

RESULT 1	
AA91945	
ID	AA91945 standard; protein; 559 AA.
AC	
XX	
XX	AA91945;
XX	
DT	19-JUL-2000 (first entry)
XX	
DE	Human chaperone protein 6 (HCHP-6).
XX	
KM	Human chaperone protein; HCHP-6; neurodegenerative; cancer; metabolic;
KW	developmental; autoimmune; inflammatory; cell proliferation.
XX	
XX	
OS	Homo sapiens.
XX	
XX	
Key	
FT	Location/Qualifiers
FT	6
FT	/note= "potential phosphorylation site"
FT	14..82
FT	/note= "DnaJ signature sequence"
FT	25..44
FT	/note= "DnaJ signature sequence"
FT	26
FT	/note= "potential phosphorylation site"
FT	53
FT	/note= "potential phosphorylation site"
FT	59..78
FT	/note= "DnaJ signature sequence"
FT	99
FT	/note= "potential phosphorylation site"
FT	173
FT	/note= "potential phosphorylation site"
FT	203
FT	/note= "potential phosphorylation site"
FT	204
FT	/note= "potential phosphorylation site"
FT	232
FT	/note= "potential phosphorylation site"
FT	246
FT	/note= "potential phosphorylation site"
FT	281
FT	/note= "potential glycosylation site"
FT	325
FT	/note= "potential phosphorylation site"
FT	333
FT	/note= "potential phosphorylation site"
FT	344
FT	/note= "potential phosphorylation site"

FT Modified-site 362 /note= "potential phosphorylation site"  
 FT Modified-site 448 /note= "potential phosphorylation site"  
 FT Modified-site 477 /note= "potential phosphorylation site"  
 FT Modified-site 480 /note= "potential phosphorylation site"  
 FT Modified-site 552 /note= "potential phosphorylation site"  
 FT Modified-site /note= "potential phosphorylation site"  
 PN W0200017358-A2.  
 PD 30-MAR-2000.  
 XX  
 XX 22-SEP-1999; 9800-US022027.  
 PF 22-SEP-1999; 9800-US022027.  
 XX  
 XX 22-SEP-1999; 9800-US022027.  
 PR 22-SEP-1999; 9800-US022027.  
 PR 19-JAN-1999; 9905-00233291.  
 PR 19-APR-1999; 9905-00294698.  
 PR 19-APR-1999; 9905-0172232P.  
 XX  
 XX (INCY-) INCYTE PHARM INC.  
 XX  
 XX Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;  
 PI WPI; 2000-283583/24.  
 XX N-PSDB; AAA08568.  
 DR  
 XX  
 XX New purified polypeptides and polynucleotides encoding human chaperone  
 PT proteins, useful for diagnosing, treating and preventing disorders  
 XX associated with the expression human chaperone proteins.  
 PS Claim 1; Page 75-76; 88pp; English.  
 XX  
 XX AAA08563-68 encode human chaperone proteins 1-6 (HCHP-1 to HCHP-6)  
 CC respectively. The sequences can be used to treat and prevent disorders  
 CC associated with altered expression or activity of HCHP comprising  
 CC administering a composition comprising the polypeptide or an antagonist  
 CC to a patient (claimed). The human chaperone proteins are also useful for  
 CC the diagnosis, treatment or prevention of neurodegenerative, metabolic,  
 CC developmental, autoimmune/inflammatory disorders and cell proliferative  
 CC disorders including cancer  
 XX  
 XX Sequence 559 AA:  
 SQ  
 Query Match 100.0%; Score 2857; DB 3; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-263;  
 Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 301 IPHSFALISYQHKFODDQTRVKGSLKAGFGTVVEYGAEKRSIRHSVILGAASVGVPOG 360  
 QY 361 VSLKVTNLSAQTTPPPHILTDOLPSAMFVTVGPIVYFPMHRLIIPYRAQKEKL 420  
 DB 361 VSLKVTNLSAQTTPPPHILTDOLPSAMFVTVGPIVYFPMHRLIIPYRAQKEKL 420  
 QY 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNMYGKEVNDKSRKS 480  
 DB 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNMYGKEVNDKSRKS 480  
 QY 481 EKVKVIVDTVPLOCLVCKSKILITEASKAGLPGFYDPCVGEERKILVLYQFPGVILQVNV 540  
 DB 481 EKVKVIVDTVPLOCLVCKSKILITEASKAGLPGFYDPCVGEERKILVLYQFPGVILQVNV 540  
 QY 541 LDSEALRIPKQSHRIDTDG 559  
 DB 541 LDSEALRIPKQSHRIDTDG 559  
 RESULT 2  
 ABG79329  
 ID ABG79329 standard; protein; 559 AA.  
 XX  
 XX ABG79329;  
 AC  
 XX  
 XX 15-NOV-2002 (first entry)  
 DT  
 XX  
 XX Human foetal cell selective cDNA encoded protein #28.  
 DE  
 XX  
 XX Human; foetal liver myeloid cell; erythroblast; foetal abnormality;  
 KW maternal blood; differential display; chromosomal abnormality;  
 KW single gene disorder; aneuploidy; nucleotide triplet expansion disorder;  
 KW trisomy 13; trisomy 21; Klinefelter syndrome; spina bifida;  
 KW sickle cell anaemia; thalassemia; Marfan syndrome;  
 KW Duchenne muscular dystrophy; cystic fibrosis; Fragile X-syndrome;  
 KW Friedreich's ataxia; myotonic dystrophy; Huntington's disease;  
 KW sex determination.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX W0200255985-A2.  
 PN  
 XX  
 XX 18-JUL-2002.  
 PD  
 XX  
 XX 01-NOV-2001; 2001MO-US045340.  
 PF  
 XX  
 XX 15-NOV-2000; 2000US-0248882P.  
 PR  
 XX  
 XX (HOPE) ROCHE DIAGNOSTICS CORP.  
 PA  
 XX  
 XX Schueler PA, Xu H, Foltz L, Wu X, Sha Y, Nagy A, Mahoney WC;  
 PI WPI; 2002-619108/66.  
 DR N-PSDB; ABS64608.  
 XX  
 XX  
 XX Detecting fetal cells in maternal blood sample, useful for diagnosing an  
 PT abnormality in a fetal cell, involves using specific nucleic acid probes  
 PT that hybridize to fetal cell associated RNAs.  
 XX  
 XX  
 XX Claim 116; Page 204-205; 215pp; English.  
 PS  
 XX  
 XX The invention relates to detecting foetal cells in maternal blood sample,  
 CC involving contacting sample with a first probe or performing expression  
 CC analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to  
 CC that obtained from mature cells to identify RNA/cDNA species useful as  
 CC probe, contacting sample with the probe and identifying if the sample  
 CC comprises a cell that comprises mRNA that hybridises to the probe. The  
 CC probes are identified by differential display analysis using mature liver  
 CC cells and foetal liver myeloid cells of less than 32 weeks of gestation.  
 CC Also included are: (1) The probe sequences (appearing as ABS64586-  
 CC ABS64618), an isolated nucleic acid molecule having a sequence which is  
 CC at least 90% identical to the probe or its complement, or having a  
 CC nucleotide sequence identical to at least 15 consecutive nucleotide

CC residues of the probe, the encoded proteins from the probe or encoding a  
CC naturally occurring allelic variant or fragment; (2) a non-mammalian host  
CC cell containing the probe. The method is useful for detecting a foetal  
CC cell (such as erythroblast or trophoblast) in a maternal blood sample,  
CC which is useful for diagnosing a chromosomal abnormality, single gene  
CC disorder or nucleotide triplet expansion in the gene, in a foetal cell.  
CC For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
CC syndrome), spina bifida, sickle cell anaemia, a thalassemia, Marfan  
CC syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
CC disease, Friedreich's ataxia, myotonic dystrophy or Huntington's  
CC disease. The probes are also useful for determination of the sex of a  
CC foetus and for detecting the presence of multiple foetuses at early  
CC stages of pregnancy. The present sequence is encoded by a foetal cell  
CC specific probe of the invention

XX Sequence 559 AA;

Query Match 100.0%; Score 2857; DB 5; Length 559;

Best Local Similarity 100.0%; Pred. No. 1.4e-263; Mismatches 0; Indels 0; Gaps 0;

Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAISEELDNEDYSLNVRREASSELKAAVRRICMLYHPDKHRDPKLSQAEPLFN 60

DB 1 MATAISEELDNEDYSLNVRREASSELKAAVRRICMLYHPDKHRDPKLSQAEPLFN 60

QY 61 LVHQAAYEVLSDPOTRAIYDIYGRKGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120

DB 61 LVHQAAYEVLSDPOTRAIYDIYGRKGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120

QY 121 ORTPKGTISVGVNATDLPFRYDEEYEDVSGSSFPQIEINMHI SOSIEADPLTDTAIL 180

DB 121 ORTPKGTISVGVNATDLPFRYDEEYEDVSGSSFPQIEINMHI SOSIEADPLTDTAIL 180

QY 181 SGSLSTONGNGGSGINFALRRVTSKAGWELFEGADLQGPFLGKLFRNLTPRCFVTYN 240

DB 181 SGSLSTONGNGGSGINFALRRVTSKAGWELFEGADLQGPFLGKLFRNLTPRCFVTYN 240

QY 241 CALQFSSRGIRPGITLVNARLNDKTVGYLQWRKCIQSANMTSYRDTKTSHFYVALQLG 300

DB 241 CALQFSSRGIRPGITLVNARLNDKTVGYLQWRKCIQSANMTSYRDTKTSHFYVALQLG 300

QY 301 IPHSFALISYOHKFODDQTRVKSGLKAGFGTVVEYGAERKISRHSVLGAASVGVPOG 360

DB 301 IPHSFALISYOHKFODDQTRVKSGLKAGFGTVVEYGAERKISRHSVLGAASVGVPOG 360

QY 361 VSLKVKLNRAASQYFFPIHLTDQLPSAMFYATVGPVVVYPMARLIIKPYLRAQKEXEL 420

DB 361 VSLKVKLNRAASQYFFPIHLTDQLPSAMFYATVGPVVVYPMARLIIKPYLRAQKEXEL 420

QY 421 EKORBSAATDVLOKKQEAESAVALMOESVRRIIIEAESRMGLIIVNAMYGFVNDKSRKS 480

DB 421 EKORBSAATDVLOKKQEAESAVALMOESVRRIIIEAESRMGLIIVNAMYGFVNDKSRKS 480

QY 481 EKVVIVDTVPLQCLVQSKLILTEASKAGLPGFDPVGVSEKMLKVLVYORGVLAHQMV 540

DB 481 EKVVIVDTVPLQCLVQSKLILTEASKAGLPGFDPVGVSEKMLKVLVYORGVLAHQMV 540

QY 541 LDSEALRIPOKSHRIDTDG 559

DB 541 LDSEALRIPOKSHRIDTDG 559

RESULT 3

AAB93115 ID AAB93115 standard; protein; 559 AA.

XX AAB93115;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:11987.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX Ep1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 11987; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dt primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

XX present invention

SO Sequence 559 AA;

Query Match 99.7%; Score 2848; DB 4; Length 559;

Best Local Similarity 99.6%; Pred. No. 1e-262; Mismatches 1; Indels 0; Gaps 0;

Matches 557; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MATAISEELDNEDYSLNVRREASSELKAAVRRICMLYHPDKHRDPKLSQAEPLFN 60

DB 1 MATAISEELDNEDYSLNVRREASSELKAAVRRICMLYHPDKHRDPKLSQAEPLFN 60

QY 61 LVHQAAYEVLSDPOTRAIYDIYGRKGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120

DB 61 LVHQAAYEVLSDPOTRAIYDIYGRKGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120

QY 121 ORTPKGTISVGVNATDLPFRYDEEYEDVSGSSFPQIEINMHI SOSIEADPLTDTAIL 180

DB 121 ORTPKGTISVGVNATDLPFRYDEEYEDVSGSSFPQIEINMHI SOSIEADPLTDTAIL 180

QY 181 SGSLSTONGNGGSGINFALRRVTSKAGWELFEGADLQGPFLGKLFRNLTPRCFVTYN 240





```

Db      5 MATALSEELDNEDYSLNVRREASSELKAAVRLCMLYHPDKHRDPDLKQASRLFN 64
Qy      61 LVHQAAYEVLSDPQTRAIYDIYKRGLEMEGWEVERRTPAEIRREPERLQREERRRLQ 120
Db      65 LVHQAAYEVLSDPQTRAIYDIYKRGLEMEGWEVERRTPAEIRREPERLQREERRRLQ 124
Qy      121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHSQSIAPLTATDTAIL 180
Db      125 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHSQSIAPLTATDTAIL 184
Qy      181 SGLSTONGGGGGINFALRRVTSAKWGELFEFGADLOGPLFGLKLFRLNLTFRCEVTNN 240
Db      185 SGLSTONGGGGGINFALRRVTSAKWGELFEFGADLOGPLFGLKLFRLNLTFRCEVTNN 244
Qy      241 CALQFSRGRIRPGITTVLARNLDKNTVGYLQW 272
Db      245 CALQFSRGRIRPGITTVLARNLDKNTVGYLQW 276

```

## RESULT 7

```

ID      ABG14697 standard; protein; 360 AA.
XX
AC      ABG14697;
XX
DT      18-FEB-2002 (first entry)
XX
DE      Novel human diagnostic protein #14688.

```

```

XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM      food supplement; medical imaging; diagnostic; genetic disorder.
XX      Homo sapiens.

```

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS78884.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20; SEQ ID NO 45056; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp://ipo.int/pub/published\_pct\_sequences

XX Sequence 360 AA;

Query Match 49.0%; Score 1401; DB 4; Length 360;  
Best Local Similarity 99.6%; Pred. No. 9, 9e-125;  
Matches 271; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MATALSEELDNEDYSLNVRREASSELKAAVRLCMLYHPDKHRDPDLKQASRLFN 60
Db      5 MATALSEELDNEDYSLNVRREASSELKAAVRLCMLYHPDKHRDPDLKQASRLFN 64
Qy      61 LVHQAAYEVLSDPQTRAIYDIYKRGLEMEGWEVERRTPAEIRREPERLQREERRRLQ 120
Db      65 LVHQAAYEVLSDPQTRAIYDIYKRGLEMEGWEVERRTPAEIRREPERLQREERRRLQ 124
Qy      121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHSQSIAPLTATDTAIL 180
Db      125 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHSQSIAPLTATDTAIL 184
Qy      181 SGLSTONGGGGGINFALRRVTSAKWGELFEFGADLOGPLFGLKLFRLNLTFRCEVTNN 240
Db      185 SGLSTONGGGGGINFALRRVTSAKWGELFEFGADLOGPLFGLKLFRLNLTFRCEVTNN 244
Qy      241 CALQFSRGRIRPGITTVLARNLDKNTVGYLQW 272
Db      245 CALQFSRGRIRPGITTVLARNLDKNTVGYLQW 276

```

## RESULT 8

ID ABB63918 standard; protein; 521 AA.

XX ABB63918;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 18546.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00641150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08021.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

PS Disclosure; SEQ ID NO 18546; 21pp + Sequence Listing; English.  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is



CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB16177-  
CC AB172072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 521 AA:  
Query Match 39.5%; Score 1128.5; DB 4; Length 521;  
Best Local Similarity 44.2%; Pred. No. 2e-98;  
Matches 233; Conservative 112; Mismatches 169; Indels 13; Gaps 6;  
QY 6 SEEBLNDYSLINVRREASSEELKAYRRLCMLYHPDKRDELPKSOAERLFLNVAQA 65  
DB 8 SDAELD-ENYVTFPLNPRDPAEQINTAYRKRQSMFHPDKLDDPSKMAEIMFNRTKRA 66  
QY 66 YEVSDDPOTRAIYDYGKGLMEGWEVERRRTPARREPERLQREBERLQQRTP 125  
DB 67 YEVSDDPOTRAIYDYGKGLMEGWEVERRRTPARREPERLQREBERLQQRTP 126  
QY 126 KGTISVGADATDLPRYDEEYEDVSGSSFPQIEINKKHISQISIRAPLTATDAILSGSL 185  
DB 127 KGTITVNMATEIFAPYD-----SEMPHVEIGSMSIASIRAPITRKDMITMSGLY 179  
QY 186 TONGNGGGINFALRYTSAGKWELEFGAGDLQGPLFGKLPNLTFRCTVTNCAIQ 245  
DB 180 TONGNGGGINFALRYTSAGKWELEFGAGDLQGPLFGKLPNLTFRCTVTNCAIQ 246  
QY 246 SSRIGRLTTLVARNDKNTVGLQWRWGIOSAMNTSIYDRTSHFTVALQIGHSF 305  
DB 237 RDQGVIPALFSTLAVQDKHTMGSLLTNAGSQSSMTQIDHSKETYSLSLVLVIGTHVY 296  
QY 306 ALIISYQKHFDDPOTRAIYDYGKGLMEGWEVERRRTPARREPERLQREBERLQQRTP 365  
DB 297 FGLSYTRKM-ENELKLTAKAVGTFFGMSGYGEKYSKTSSTAYVIGVPGVILKF 355  
QY 366 KLNASQTYPPPIHLTDQLPSAMFYATVGPVVYFAMHLLIIRYLAQKEKELEKQRE 425  
DB 356 KILRSNGSYVPIHLSDEIVPAVFAVSVTVIAMFFIKRTVMDPMEAKRNIEVERTKR 415  
QY 426 SAARDVLOKQKQASAVRLMGESEVRRIIEAESRMGLIIVNATGKFPVNDKSRSEKVK 485  
DB 416 QNEORLSAKRHEASAAVHLMQATNRIIMTEELARNGLIVTAAVVGCTLAG-GRQFKPEQS 474  
QY 486 IDVTVPLOCLVDSKLLITLTKASKAGLPGFYDPVCGEENKLVLYQFR 532  
DB 475 LDVTAIQCWVKNGTLQLHDSKSDLPGFYDPGIGEDKILRIETTYQ 521

RESULT 9  
ABG03199  
ID ABG03199 standard; protein, 344 AA.  
XX  
AC ABG03199;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #3190.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN M0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX

PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS67386.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 33556; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 344 AA:  
Query Match 39.3%; Score 1124; DB 4; Length 344;  
Best Local Similarity 84.1%; Pred. No. 2.7e-98;  
Matches 232; Conservative 3; Mismatches 21; Indels 20; Gaps 4;  
QY 1 MATALSEEBLNDYSLINVRREASSEELKAYRRLCMLYHPDKRDELPKSOAERL 60  
DB 1 MATALSEEBLNDYSLINVRREASSEELKAYRRLCMLYHPDKRDELPKSOAERL 60  
QY 61 LVHQAIEVLSDPOTRAIYDYGKGLMEGWEVERRRTPARREPERLQREBERL 116  
DB 56 CL-----VTPKPPSMIYMGREBWKMDGLMWKGEP-----LEIRREFEFLRERRE 104  
QY 117 RRLQORTNPKGTISVGADATDLPRYDEEYEDVSGSSFPQIEINKKHISQISIRAPLTATD 176  
DB 105 RRLQORTNPKGTISVGADATDLPRYDEEYEDVSGSSFPQIEINKKHISQISIRAPLTATD 164  
QY 177 TAILSGSLTONGNGGGINFALRYTSAGKWELEFGAGDLQGPLFGKLPNLTFRCTVTN 236  
DB 165 TAILSGSLTONGNGGGINFALRYTSAGKWELEFGAGDLQGPLFGKLPNLTFRCTVTN 224  
QY 237 VTTNCAIQFSSRGIRGLTTLVARNDKNTVGLQW 272  
DB 225 VTTNCAIQFSSRGIRGLTTLVARNDKNTVGLQW 260

RESULT 10  
AAU17215  
ID AAU17215 standard; protein, 159 AA.  
XX  
AC AAU17215;  
XX  
DT 07-NOV-2001 (first entry)  
XX

XX DE Novel signal transduction pathway protein, Seg ID 780.  
XX  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antineoplastic; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX  
OS Homo sapiens.  
XX  
PN WO200154733-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001312.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 06-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234774P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234984P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235634P.  
PR 27-SEP-2000; 2000US-0235635P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.

PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465460/50.  
DR N-PSDB; AAS27132.  
XX  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT diagnosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders.  
XX  
PS Claim 1; SEQ ID NO 780; 880bp; English.  
XX  
CC The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative disorders  
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.  
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities  
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.  
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),  
CC respiratory disorders, dermatological disorders, in wound healing,  
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's  
CC disease), reproductive system disorders, gastrointestinal disorder  
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of  
CC B-cell responsiveness to pathogens, activators of T-cells, to induce  
CC higher affinity antibodies, and as a means to induce tumour proliferation  
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-  
CC AAU17683 represent novel signal transduction pathway protein, amino acid  
XX sequences of the invention

Query Match 26.3%; Score 750; DB 4; Length 159;  
Best Local Similarity 92.5%; Pred. No. 4.4e-63;  
Matches 147; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 2 ATATSEELNEDYVSLNVRREASSELKAAVRLCMLYHPDGRPELKSQARLNL 61  
DB 1 ATATSEELNEDYVSLNVRREASSELKAAVRLCMLYHPDGRPELKSQARLNL 60  
QY 62 VHQAYEVLSDPQTALYDIYKRGLEMGWEVVERRRTPARIEEFERLQREERRRLOQ 121  
DB 61 VHQAYEVLSDPQTALYDIYKRGLEMGWEVVERRRTPARIEEFERLQREERRRLOQ 120  
QY 122 RTNPKGTISVGVNATDLFDRYDEEYEDVSGSSFPQIEIN 160  
DB 121 RTNPKGTISVGVNATDLFDRYDXLXKMCAPAVXFPXIEIN 159

RESULT 11  
ADB93923  
ID ADB93923 standard; protein, 159 AA.  
XX  
AC ADB93923;  
XX

DT 04-DEC-2003 (first entry)  
XX  
DE Human novel protein #157.  
XX  
KW human; autoimmune disease; Parkinson's disease; silicosis;  
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;  
KW immunosuppressive agent; adjuvant; enhance immune response;  
KW higher affinity antibody induction;  
KW increased serum immunoglobulin concentration.  
XX  
OS Homo sapiens.  
XX  
PN US2002168711-A1.  
XX  
PD 14-NOV-2002.  
XX  
PF 17-JAN-2001; 2001US-00764868.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180688P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218280P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBEN/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX

PI Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-719985/68.  
DR N-PSDB; ADB93300.  
XX  
PT New isolated polypeptide useful for diagnosing and treating  
PT immunosuppressive conditions such as autoimmune disease and Parkinson's  
PT disease.  
PS  
PS Claim 11: SEQ ID NO 780; 345bp; English.  
XX  
XX The invention relates to an isolated polypeptide. The polypeptide is  
CC useful for diagnosing a pathological condition or a susceptibility to a  
CC pathological condition in a subject, by determining the presence or  
CC amount of expression of the polypeptide in a biological sample and  
CC diagnosing a pathological condition or a susceptibility to a pathological  
CC condition based on the presence or amount of expression of the  
CC polypeptide. The polypeptide is also useful for identifying a binding  
CC partner to the polypeptide, which involves contacting the polypeptide  
CC with a binding partner and determining whether the binding partner  
CC effects an activity of the polypeptide. The polypeptide or the nucleic  
CC acid encoding the polypeptide is useful for preventing, treating, or  
CC ameliorating a medical condition, which involves administering the  
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid  
CC is useful for diagnosing a pathological condition or a susceptibility to  
CC a pathological condition in a subject, which involves determining the  
CC presence or absence of a mutation in the nucleic acid, and diagnosing a  
CC pathological condition or susceptibility to a pathological condition  
CC based on the presence or absence of the mutation. The polypeptide, the  
CC nucleic acid and an antibody to the polypeptide are useful for treating  
CC autoimmune disease, Parkinson's disease, sickle cell disease, Crohn's  
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,  
CC the nucleic acid and the antibody are useful as immunosuppressive agents,  
CC as adjuvants to enhance immune responses, and as agents to induce higher  
CC affinity antibodies and increase serum immunoglobulin concentrations. The  
CC present sequence represents the amino acid sequence of a novel human  
CC protein. Note: The sequence data for this patent did not form part of the  
CC printed specification but was obtained in electronic format direct from  
CC USPTO at [seqdata.uspto.gov/sequence.html?docid=20020168711](http://seqdata.uspto.gov/sequence.html?docid=20020168711).  
XX  
XX Sequence 159 AA:  
SO  
Query Match 26.3%; Score 750; DB 7; Length 159;  
Best Local Similarity 92.5%; Pred. No. 4,4e-63;  
Matches 147; Conservative 1; Mismatches 11; Indels 0; Gaps 0;  
QY 2 ATATSEELNEDYYSILNRRRASSSEELKAAVRLCMLYHPKRDPELKSQAERLFL 61  
DB 1 ATATSEELNEDYYSILNRRRASSSEELKAAVRLCMLYHPKRDPELKSQAERLFL 60  
QY 62 VHQAYEVLSDPQTRATIDYGRGLMEGMEVVERRTPTAIEEPEERLQREERERLQ 121  
DB 61 VHQAYEVLSDPQTRATIDYGRGLMEGMEVVERRTPTAIEEPEERLQREERERLQ 120  
QY 122 RINPKGTISGVVDATDLFDRIYDEYEDVSGSSPFOIEIN 160  
DB 121 RINPKGTISGVVDATDLFDRIYDXLKMCPAVXFXPIEIN 159  
RESULT 12  
AAU17589  
ID AAU17589 standard; protein; 159 AA.  
XX  
XX AAU17589;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DE Novel signal transduction pathway protein, Seq ID 1154.  
XX  
XX  
XX Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;

KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200154733-A1.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001312.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225476P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 14-AUG-2000; 2000US-0225797P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226682P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0226924P.  
PR 01-SEP-2000; 2000US-0229387P.  
PR 01-SEP-2000; 2000US-0229387P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0233968P.  
PR 14-SEP-2000; 2000US-0233297P.  
PR 14-SEP-2000; 2000US-0233298P.  
PR 14-SEP-2000; 2000US-0233299P.  
PR 14-SEP-2000; 2000US-0233400P.  
PR 14-SEP-2000; 2000US-0233401P.  
PR 14-SEP-2000; 2000US-0233063P.

PR	14-SEP-2000;	2000US-0231865P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235836P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236369P.
PT	02-OCT-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	01-NOV-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	17-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250316P.
PR	01-DEC-2000;	2000US-0250317P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0231865P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235836P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236369P.
PT	02-OCT-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-O	

KM higher affinity antibody induction;  
XX increased serum immunoglobulin concentration.  
OS Homo sapiens.  
XX US2002168711-A1.  
XX PD 14-NOV-2002.  
XX PF 17-JAN-2001; 2001US-00764868.  
XX 31-JAN-2000; 2000US-0179065P.  
XX PR 04-FEB-2000; 2000US-0180628P.  
XX PR 28-JUN-2000; 2000US-0214886P.  
XX PR 07-JUL-2000; 2000US-0216647P.  
XX PR 07-JUL-2000; 2000US-0216880P.  
XX PR 11-JUL-2000; 2000US-0217487P.  
XX PR 11-JUL-2000; 2000US-0217496P.  
XX PR 14-JUL-2000; 2000US-0218290P.  
XX PR 26-JUL-2000; 2000US-0220963P.  
XX PR 26-JUL-2000; 2000US-0220964P.  
XX PR 14-AUG-2000; 2000US-0224518P.  
XX PR 14-AUG-2000; 2000US-0224519P.  
XX PR 14-AUG-2000; 2000US-0225267P.  
XX PR 14-AUG-2000; 2000US-0225268P.  
XX PR 14-AUG-2000; 2000US-0225270P.  
XX PR 14-AUG-2000; 2000US-0225447P.  
XX PR 14-AUG-2000; 2000US-0225757P.  
XX PR 14-AUG-2000; 2000US-0225758P.  
XX PR 22-AUG-2000; 2000US-0226868P.  
XX PR 30-AUG-2000; 2000US-0228284P.  
XX PR 01-SEP-2000; 2000US-0229287P.  
XX PR 01-SEP-2000; 2000US-0229343P.  
XX PR 01-SEP-2000; 2000US-0229344P.  
XX PR 01-SEP-2000; 2000US-0229345P.  
XX PR 05-SEP-2000; 2000US-0229509P.  
XX PR 05-SEP-2000; 2000US-0229513P.  
XX PR 08-SEP-2000; 2000US-0231413P.  
XX PR 21-SEP-2000; 2000US-0234223P.  
XX PR 21-SEP-2000; 2000US-0234274P.  
XX PR 25-SEP-2000; 2000US-0234997P.  
XX PR 27-SEP-2000; 2000US-0235834P.  
XX PR 29-SEP-2000; 2000US-0236337P.  
XX PR 29-SEP-2000; 2000US-0236337P.  
XX PR 29-SEP-2000; 2000US-0236367P.  
XX PR 29-SEP-2000; 2000US-0236368P.  
XX PR 29-SEP-2000; 2000US-0236369P.  
XX PR 29-SEP-2000; 2000US-0236370P.  
XX PR 02-OCT-2000; 2000US-0236802P.  
XX PR 02-OCT-2000; 2000US-0237037P.  
XX PR 02-OCT-2000; 2000US-0237038P.  
XX PR 02-OCT-2000; 2000US-0237039P.  
XX PR 02-OCT-2000; 2000US-0237040P.  
XX PR 13-OCT-2000; 2000US-0239935P.  
XX PR 20-OCT-2000; 2000US-0240960P.  
XX PR 20-OCT-2000; 2000US-0241785P.  
XX PR 20-OCT-2000; 2000US-0241809P.  
XX PR 01-NOV-2000; 2000US-0244617P.  
XX PR 17-NOV-2000; 2000US-0249299P.  
XX PR 08-DEC-2000; 2000US-0251856P.  
XX PR 08-DEC-2000; 2000US-0251868P.  
XX PR 08-DEC-2000; 2000US-0251869P.  
XX (ROSE/) ROSEN C A.  
XX (RUBEN/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX PI Rosen CA, Ruben SM, Barash SC;  
XX WPI: 2003-719985/68.  
XX DR N-PSDB; ADB93674.  
XX PT New isolated polypeptide useful for diagnosing and treating  
PT immunosuppressive conditions such as autoimmune disease and Parkinson's

PT disease.  
XX Claim 11; SEQ ID NO 1154; 345pp; English.  
XX CC The invention relates to an isolated polypeptide. The polypeptide is  
XX CC useful for diagnosing a pathological condition or a susceptibility to a  
XX CC pathological condition in a subject, by determining the presence or  
XX CC amount of expression of the polypeptide in a biological sample and  
XX CC diagnosing a pathological condition or a susceptibility to a pathological  
XX CC condition based on the presence or amount of expression of the  
XX CC polypeptide. The polypeptide is also useful for identifying a binding  
XX CC partner to the polypeptide, which involves contacting the polypeptide  
XX CC with a binding partner and determining whether the binding partner  
XX CC affects an activity of the polypeptide. The polypeptide or the nucleic  
XX CC acid encoding the polypeptide is useful for preventing, treating, or  
XX CC ameliorating a medical condition, which involves administering the  
XX CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid  
XX CC is useful for diagnosing a pathological condition or a susceptibility to  
XX CC a pathological condition in a subject, which involves determining the  
XX CC presence or absence of a mutation in the nucleic acid, and diagnosing a  
XX CC pathological condition or susceptibility to a pathological condition  
XX CC based on the presence or absence of the mutation. The polypeptide, the  
XX CC nucleic acid and an antibody to the polypeptide are useful for treating  
XX CC autoimmune disease, Parkinson's disease, sickle cell disease, Crohn's  
XX CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,  
XX CC the nucleic acid and the antibody are useful as immunosuppressive agents,  
XX CC as adjuvants to enhance immune responses, and as agents to induce higher  
XX CC affinity antibodies and increase serum immunoglobulin concentrations. The  
XX CC present sequence represents the amino acid sequence of a novel human  
XX CC protein. Note: The sequence data for this patent did not form part of the  
XX CC printed specification but was obtained in electronic format direct from  
XX CC USPTO at [seqdata.uspto.gov/sequence.html?DocId=20020168711](http://seqdata.uspto.gov/sequence.html?DocId=20020168711).  
XX SEQ Sequence 159 AA;  
XX Query Match 25.7%; Score 733; DB 7; Length 159;  
XX Best Local Similarity 90.6%; Pred. No. 1.9e-61;  
XX Matches 144; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
XX QY 2 ATRALSEELDNEDDYSLNVRREASSEELKAYRRLCMYHPDKRDPDLKQARLFTL 61  
XX Db 1 ATRALSEELDNEDDYSLNVRREASSEELKAYRRLCMYHPDKRDPDLKQARLXNL 60  
XX QY 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAEIREEFERLQREERRRLQ 121  
XX Db 61 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAEIREEFERLQREERRRLQ 120  
XX QY 122 RTNPKGTISVGDATDLFDRYDEYDVSGSSFPQIEIN 160  
XX Db 121 RTNPKGTISVGDATDLFDRYDXLXLMCPAVXFPXIEIN 159  
XX RESULT 14  
XX ABG14695  
XX ID ABG14695 standard; protein; 147 AA.  
XX AC ABG14695;  
XX XX 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #14686.  
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX XX WO200175067-A2.  
XX XX 11-OCT-2001.  
XX XX 30-MAR-2001; 2001WO-US008631.  
XX PF

PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX  
 PI WPI; 2001-639362/73.  
 XX N-PSDB; AAS78882.  
 DR  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 45054; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridization probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 147 AA;  
 Query Match 23.7%; Score 676.5; DB 4; Length 147;  
 Best Local Similarity 84.0%; Pred. No. 4.1e-56;  
 Matches 137; Conservative 0; Mismatches 1; Indels 25; Gaps 1;  
 QY 4 ALSEBEILDNDYSLNVRREASSELKAYRLCMLYHPDKRDPKLSQASRLFN 63  
 DB 10 ALSEBEILDNDYSLNVRREASSELKAYRLCMLYHPDKRDPKLSQASRLFN 69  
 QY 64 QAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 123  
 DB 70 QAY-----EYERRRTPAEIRREFFERLQRRERRLQORT 104  
 QY 124 NPKGTISGVNATDLPDYDEYEDVSGSFPQIENKMHISQ 166  
 DB 105 NPKGTISGVNATDLPDYDEYEDVSGSFPQIENKMHISQ 147  
 RESULT 15  
 ID AAG01815  
 XX AAG01815 standard; protein; 119 AA.  
 AC AAG01815;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 5896.  
 XX  
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX

PN EP1033401-A2.  
 PD 06-SEP-2000.  
 XX  
 PD 21-FEB-2000; 2000EP-00200610.  
 XX  
 PF 26-FEB-1999; 99US-0122487P.  
 PR  
 XX (BEST) GENSET.  
 PA  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 XX WPI; 2000-500381/45.  
 DR N-PSDB; AAC01821.  
 DR  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 XX  
 PS Claim 13; SEQ ID NO 5896; 71pp + Sequence Listing; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number of  
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
 CC sequences derived from the 5' ends of mRNAs and even in those cases where  
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
 CC are also used in diagnostic, forensic, gene therapy and chromosome  
 CC mapping procedures. They are used to obtain upstream regulatory sequences  
 CC and to design expression and secretion vectors  
 CC  
 XX  
 SQ Sequence 119 AA;  
 Query Match 21.5%; Score 613; DB 3; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-50;  
 Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATALSEBEILDNDYSLNVRREASSELKAYRLCMLYHPDKRDPKLSQASRLFN 60  
 DB 1 MATALSEBEILDNDYSLNVRREASSELKAYRLCMLYHPDKRDPKLSQASRLFN 60  
 QY 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 119  
 DB 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 119  
 Search completed: July 13, 2004, 08:40:48  
 Job time : 86 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 24 Seconds  
(without alignments)  
1202.455 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATALSEEDLNEDYSLN.....VLDSALRIPKSHRIDTDG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	7.8	330	2	US-08-868-288A-3
2	223	7.8	330	3	US-09-235-373-3
3	223	7.8	330	3	US-09-388-993-3
4	195	6.8	392	4	US-09-198-452A-48
5	194	6.8	340	2	US-08-974-546-5
6	192	6.7	337	4	US-09-665-479A-8
7	192	6.6	348	4	US-08-974-546-1
8	189	6.6	87	4	US-09-882-835-4
9	187	6.5	277	2	US-08-868-288A-7
10	187	6.5	277	3	US-09-235-373-7
11	187	6.5	277	3	US-09-388-993-7
12	187	6.5	351	2	US-08-868-288A-6
13	187	6.5	351	3	US-09-235-373-6
14	187	6.5	351	3	US-09-388-993-6
15	186	6.5	153	4	US-09-621-976-5129
16	185	6.5	381	4	US-09-253-991A-27174
17	181	6.3	380	4	US-09-489-039A-9011
18	180.5	6.3	419	2	US-08-686-417-3
19	177.5	6.2	399	4	US-09-553-498-2
20	177.5	6.2	399	4	US-09-618-869-2
21	176.5	6.2	380	4	US-09-543-681A-5912
22	175.5	6.1	131	4	US-09-553-498-4
23	175.5	6.1	131	4	US-09-618-869-4
24	171.5	6.0	375	4	US-09-328-352-4984
25	169	5.9	352	4	US-08-472-534-6
26	166	5.8	127	4	US-09-370-838-199
27	166	5.8	397	2	US-08-868-288A-5

28	166	5.8	397	3	US-09-235-373-5	Sequence 5, Appli
29	166	5.8	397	3	US-09-388-993-5	Sequence 5, Appli
30	164.5	5.8	358	2	US-08-868-288A-1	Sequence 1, Appli
31	164.5	5.8	358	3	US-09-235-373-1	Sequence 1, Appli
32	164.5	5.8	358	3	US-09-388-993-1	Sequence 1, Appli
33	161.5	5.7	385	4	US-09-134-001C-3688	Sequence 3688, Ap
34	160.5	5.6	332	4	US-09-882-835-2	Sequence 2, Appli
35	160.5	5.6	407	4	US-09-540-236-2899	Sequence 2899, Ap
36	158.5	5.5	320	4	US-09-328-352-7937	Sequence 7937, Ap
37	156	5.5	438	2	US-08-897-340-34	Sequence 34, Appli
38	156	5.5	438	3	US-09-252-329-34	Sequence 34, Appli
39	156	5.5	484	2	US-08-879-260-4	Sequence 4, Appli
40	156	5.5	484	3	US-09-231-529-4	Sequence 4, Appli
41	156	5.5	484	3	US-08-977-816-4	Sequence 4, Appli
42	153	5.4	223	4	US-09-658-544-4	Sequence 4, Appli
43	151	5.3	504	3	US-09-231-529-6	Sequence 6, Appli
44	151	5.3	504	3	US-08-977-816-6	Sequence 6, Appli
45	150.5	5.3	320	4	US-09-543-681A-7884	Sequence 7884, Ap

## ALIGNMENTS

RESULT 1  
US-08-868-288A-3

Sequence 3, Application US/08868288A

Patent No. 5922567

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Lal, Preeti

APPLICANT: Bandman, Olga

TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSER: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/868,288A

FILING DATE: June 3, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 330 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: HNTZPAT01

CLONE: 260873

US-08-868-288A-3

Query Match 7.8%; Score 223; DB 2; Length 330;  
Best Local Similarity 32.1%; Pred. No. 1.8e-14;  
Matches 67; Conservative 22; Mismatches 50; Indels 7;

```

Qy 14 DYVLLAVREASASEELKAAYYRLCMYYHDDKHDELDSQAARLPLNVAQAYEVLSDPQ 73
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3 DYEVLVQVHASPSEDITKKAYRKALAMHPDK--NPNKNEAARKFKQVAEAYEVLSDAK 60
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 74 TRAIYDIYGRKGLMEG-----WEVVERRRTPAIREFEFRLQREERERRLQORTN 12
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 61 KRDIYDKYKGEGLNCGGGGSHFDSPEPEFTFRNDDVYRE----- 10
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 125 PKGTISGVDAITDLPDYDEEYEDVSG-----SSFPQLEINKQHI 16
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 104 -----PGGRDPFSGDFDFDFPFDFPFGRNRGRGRSRGTGSAFSAFSGFP----- 14
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 165 SQSIEAPLPTATPAILIS--GSLSTQNGNCGG 192
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 149 --SFGSGPSSFDGTGFTSPGSL---GHGG 171
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 2  
US-09-235-373-3  
; Sequence 3, Application US/09235373  
; Patent No. 6001509

```

1 GENERAL INFORMATION:
2 APPLICANT: Au-Young, Janice
3 APPLICANT: Lal, Preeti
4 APPLICANT: Bandman, Olga
5 TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
6 NUMBER OF SEQUENCES: 7
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Incyte Pharmaceuticals, Inc.
9 STREET: 3174 Porter Drive
10 CITY: Palo Alto
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94304
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Diskette
16 COMPUTER: IBM Compatible
17 OPERATING SYSTEM: DOS
18 SOFTWARE: FastSeq for Windows Version 2.0
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/235,373
21 FILING DATE:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 08/868,288
24 FILING DATE: June 3, 1997
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Billings, Lucy J.
27 REGISTRATION NUMBER: 36,749
28 REFERENCE/DOCKET NUMBER: PF-0309 US
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-855-0555
31 TELEFAX: 415-845-4166
32 INFORMATION FOR SEQ ID NO: 3:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 330 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 LIBRARY: HNTZPAT01
40 CLONE: 260873
41
42 JS-09-235-373-3

```

Query Match	7.8%;	Score 223;	DB 3;	Length 330;
-------------	-------	------------	-------	-------------

Matches 67; Conservative 22; Mismatches 50; Indels 70; Gaps 7;

Oy 14 DYSSLNLRREASSSEELKAAYYRLCMLYHEDKDRPELPSKQAERLFNLVHQAIEYVLSDDQ 73  
Db 3 DYEEVLGVORHASPEDIDKAAYYRKALKWHEDK--NPENKKEAEERFKQVAEEAYEYVLSDAK 60

```

QY      74 TRAIYDIYKRGIMEG-----WEVERRRTPAEIRREERLÖRRERRLQÖRTN 124
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

**D3** 6 KRDIDYKXKEGLNGGGGGSHFSDSPFEGFTFRNPDPVFERF-----103  
**QY** 125 PKRTISVGDADLDPRVDEYEDVSG-----SSPQIENGMHI 164  
       :::|::|  
**D3** 104 -----FGGRDPSPFDPEEDPEDFGNRRGPRGSRNGTGSFSAFGFP-----148  
**QY** 165 SSGIEAPLTATDTAILS--GSLSGTONGNG 192  
       :::|::|  
**D3** 149 --SFGSGFSSFDIGFTSGSL-----GHGG 171

RESULT 3

; Sequence 3, Application US/09388993

```

1 GENERAL INFORMATION:
2 APPLICANT: Au-Young, Janice
3 APPLICANT: Lal, Preeti
4 APPLICANT: Bandman, Olga
5 TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
6 NUMBER OF SEQUENCES: 7
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Incyte Pharmaceuticals, Inc.
9 STREET: 3174 Porter Drive
10 CITY: Palo Alto
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94304
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Diskette
16 COMPUTER: IBM Compatible
17 OPERATING SYSTEM: DOS
18 SOFTWARE: FastSeq for Windows Version 2.0
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/388,993
21 FILING DATE:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 08/868,288
24 FILING DATE:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Billings, Lucy J.
27 REGISTRATION NUMBER: 36,749
28 REFERENCE/DOCKET NUMBER: PF-0309 US
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-855-0555
31 TELEFAX: 415-845-4166
32 INFORMATION FOR SEQ ID NO: 3:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 330 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 LIBRARY: HNT2RAT01
40 CLONE: 260873
41 JS-09-388-993-3

```

Query Match	7.8%	Score 223;	DB 3;	Length 330;

Matches 67; Conservative 22; Mismatches 50; Indels 70; Gaps 7;

QY 14 DYSSLNVRREASSEELKAYRRLCMLYHPDKHRDPELKSQAERLFNLVHOAYEVLSDPQ 73

Db	3	DYEEVLGVRHASPEDIKAYRKALKWHPDK--NPENKEAEARKQVAAYEVLSDAK	60
Gy	74	TRAIYDIYCKRGILEMEG-----WEVVERRTPAETIEEFERLQREBERRLQORTN	124

Db 61 KRDIYDKYGKEGI.NGGGGGGSHFDSPEEFEGTFRNPDDVERFE----- 103

QY 125 PKGTISVGVDATDLFDRYDEEYEDVSG-----SSFPQIEINKMHI 164

```

Db      104 -----FGGRDPFSDFEDPFEGNRGPGSGRSRGTSFFSAFSGFP----- 148

```



```

? INFORMATION FOR SEQ ID NO: 7:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 277 amino acids
?
? TYPE: amino acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? IMMEDIATE SOURCE:
?
? LIBRARY: GenBank

```









***This Page Blank (usp/cj)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:41:11 ; Search time 352 Seconds  
(without alignments)  
495.347 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATLSEEDLNEDYSLN.....VLDSALRIRKSHRIDTDG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 1279676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCRT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCRTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2857	100.0	559	US-10-000-897-70	Sequence 70, App1
2	2857	100.0	559	US-10-408-765A-1911	Sequence 911, Ap
3	2848	99.7	559	US-10-408-765A-941	Sequence 941, App1
4	1723	60.3	336	US-10-000-897-69	Sequence 69, App1
5	750	26.3	159	US-09-764-868-780	Sequence 780, App
6	733	25.3	159	US-09-764-868-1154	Sequence 1154, Ap
7	661	23.1	569	US-10-437-963-131723	Sequence 131723,
8	456.5	16.0	118	US-09-927-738-17	Sequence 17, App1
9	451	15.8	103	US-10-000-897-71	Sequence 71, App1
10	364	12.7	338	US-10-424-599-225707	Sequence 225707,
11	314	11.0	234	US-10-424-599-184546	Sequence 184546,
12	300.5	10.5	322	US-10-425-114-71459	Sequence 71459, A
13	239.5	8.4	369	US-10-369-493-2975	Sequence 2975, Ap
14	223	7.8	242	US-10-264-049-2674	Sequence 2674, Ap
15	222.5	7.8	376	US-10-369-493-47	Sequence 47, App1

16	220.5	7.7	342	US-10-437-963-126854	Sequence 126854,
17	217	7.6	158	US-10-425-114-48912	Sequence 48912, A
18	213	7.5	116	US-10-424-599-233524	Sequence 233524,
19	210.5	7.4	340	US-10-424-599-278129	Sequence 278129,
20	210.5	7.4	362	US-10-437-963-190831	Sequence 190831,
21	210	7.4	146	US-10-424-599-191916	Sequence 191916,
22	208	7.3	153	US-10-424-599-257493	Sequence 257493,
23	206	7.2	404	US-10-425-114-59964	Sequence 59964, A
24	205.5	7.2	356	US-10-437-963-126176	Sequence 126176,
25	204	7.1	192	US-10-424-599-237535	Sequence 237535,
26	203	7.1	545	US-10-437-963-114080	Sequence 114080,
27	202	7.1	217	US-10-424-599-257471	Sequence 257471,
28	198	6.9	411	US-10-425-114-46080	Sequence 46080, A
29	197	6.9	289	US-10-424-599-192793	Sequence 192793,
30	196	6.9	352	US-10-425-114-37469	Sequence 37469, A
31	195.5	6.8	276	US-10-424-599-191915	Sequence 191915,
32	195	6.8	392	US-10-289-762-48	Sequence 48, App1
33	194	6.8	340	US-10-424-599-253715	Sequence 253715,
34	194	6.8	340	US-10-116-275-134	Sequence 134, App
35	194	6.8	375	US-10-369-493-18988	Sequence 18988, A
36	194	6.8	433	US-10-425-114-60908	Sequence 60908, A
37	193	6.8	383	US-10-437-963-108162	Sequence 108162,
38	193	6.8	440	US-10-424-599-253715	Sequence 253715,
39	193	6.8	478	US-10-425-114-65215	Sequence 65215, A
40	192	6.7	433	US-10-425-114-63266	Sequence 63266, A
41	192	6.7	435	US-10-425-114-59462	Sequence 59462, A
42	191	6.7	288	US-10-369-493-8874	Sequence 8874, Ap
43	191	6.7	334	US-10-161-051-134	Sequence 134, App
44	191	6.7	435	US-10-425-114-65366	Sequence 65366, A
45	190	6.7	365	US-10-369-493-13503	Sequence 13503, A

ALIGNMENTS

RESULT 1									
US-10-000-897-70									
Sequence 70, Application US/10000897									
Publication No. US20030165852A1									
GENERAL INFORMATION:									
APPLICANT: Schueler, P.									
TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE									
FILE REFERENCE: 11012-004-999									
CURRENT APPLICATION NUMBER: US/10/000,897									
CURRENT FILING DATE: 2001-11-15									
PRIOR APPLICATION NUMBER: 60/248,882									
PRIOR FILING DATE: 2000-11-15									
NUMBER OF SEQ ID NOS: 78									
SOFTWARE: PatentIn version 3.0									
SEQ ID NO 70									
LENGTH: 559									
TYPE: PRT									
ORGANISM: Homo sapiens									
US-10-000-897-70									
Query Match									
Best Local Similarity 100.0%; Score 2857; DB 14; Length 559;									
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	MATALSEEDLNEDDYSLNVRREASSSEELKKAAYRRLCMLYHPDGHDRDELSQASERLFN	60						
Db	1	MATRLSEEDLNEDDYSLNVRREASSSEELKKAAYRRLCMLYHPDGHDRDELSQASERLFN	60						
Qy	61	LVAQAEVLSDDPQTRAIYDIYKRGGLMEGWEVERRRTPAIREEPERLQREERRRLO	120						
Db	61	LVAQAEVLSDDPQTRAIYDIYKRGGLMEGWEVERRRTPAIREEPERLQREERRRLO	120						
Qy	121	QRNPNGGTSVGVADATDLPDRYDEEYEDVSSSPFOIEINKHISQSTIEAPLTATDTAIL	180						
Db	121	QRNPNGGTSVGVADATDLPDRYDEEYEDVSSSPFOIEINKHISQSTIEAPLTATDTAIL	180						
Qy	181	SGSLSTONGNGGGSINFALRRVTSAKMGWELEFGAGDLOGPLFGKLFRLNLTPRCFVTYN	240						

```
Db 181 SGLSTONGNGGSSINPALRRVTSAGKMGELFAGDLOGPLFGKLPRLTFRCEVTNN 240
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
Qy 301 IPHSFALISYQHKFODDDQTRVKGSLKAGFGTVVEYGAEKISRSISVLGAASVGVPOG 360
Db 301 IPHSFALISYQHKFODDDQTRVKGSLKAGFGTVVEYGAEKISRSISVLGAASVGVPOG 360
Qy 361 VSLKVLNRAASQTYFFPHLTDLPLSAMFYATVPLVYFAMHRLIIKPYLRADQEKEL 420
Db 361 VSLKVLNRAASQTYFFPHLTDLPLSAMFYATVPLVYFAMHRLIIKPYLRADQEKEL 420
Qy 421 EKQRESAATDVLOKQKQAEASAVRLMOESVRIIEAESRMGLIYNAMYGKFPVNDKSRKS 480
Db 421 EKQRESAATDVLOKQKQAEASAVRLMOESVRIIEAESRMGLIYNAMYGKFPVNDKSRKS 480
Qy 481 EKQKVIDVTVPLQCLVKOSKILITBASKAGLPGFYDPCVGEENKLVLYQFRGVLHQVWV 540
Db 481 EKQKVIDVTVPLQCLVKOSKILITBASKAGLPGFYDPCVGEENKLVLYQFRGVLHQVWV 540
Qy 541 LDSEALRIPKQSHRIDTDG 559
Db 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 2

```
US-10-408-765A-1911
/ Sequence 1911, Application US/10408765A
/ Publication No. US20040101874A1
/ GENERAL INFORMATION:
/ APPLICANT: Ghosh, Soumitra S.
/ APPLICANT: Fahy, Bojin D.
/ APPLICANT: Zhang, Bing
/ APPLICANT: Gibson, Bradford W.
/ APPLICANT: Taylor, Steven W.
/ APPLICANT: Glenn, Gary M.
/ APPLICANT: Warnock, Dale E.
/ TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
/ FILE REFERENCE: 660088.465
/ CURRENT APPLICATION NUMBER: US/10/408,765A
/ NUMBER OF SEQ ID NOS: 3077
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1911
/ LENGTH: 559
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-408-765A-1911
```

```
Query Match 100.0%; Score 2857; DB 16; Length 559;
Best Local Similarity 100.0%; Pred. No. 3.9e-263;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MATALSEEBELNDYYSILNVRREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Db 1 MATALSEEBELNDYYSILNVRREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Qy 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREFFERLOREBERRLQ 120
Db 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREFFERLOREBERRLQ 120
Qy 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIENKMHISQSIEAPLTATDTAIL 180
Db 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIENKMHISQSIEAPLTATDTAIL 180
Qy 181 SGLSTONGNGGSSINPALRRVTSAGKMGELFAGDLOGPLFGKLPRLTFRCEVTNN 240
Db 181 SGLSTONGNGGSSINPALRRVTSAGKMGELFAGDLOGPLFGKLPRLTFRCEVTNN 240
```

```
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
Qy 301 IPHSFALISYQHKFODDDQTRVKGSLKAGFGTVVEYGAEKISRSISVLGAASVGVPOG 360
Db 301 IPHSFALISYQHKFODDDQTRVKGSLKAGFGTVVEYGAEKISRSISVLGAASVGVPOG 360
Qy 361 VSLKVLNRAASQTYFFPHLTDLPLSAMFYATVPLVYFAMHRLIIKPYLRADQEKEL 420
Db 361 VSLKVLNRAASQTYFFPHLTDLPLSAMFYATVPLVYFAMHRLIIKPYLRADQEKEL 420
Qy 421 EKQRESAATDVLOKQKQAEASAVRLMOESVRIIEAESRMGLIYNAMYGKFPVNDKSRKS 480
Db 421 EKQRESAATDVLOKQKQAEASAVRLMOESVRIIEAESRMGLIYNAMYGKFPVNDKSRKS 480
Qy 481 EKQKVIDVTVPLQCLVKOSKILITBASKAGLPGFYDPCVGEENKLVLYQFRGVLHQVWV 540
Db 481 EKQKVIDVTVPLQCLVKOSKILITBASKAGLPGFYDPCVGEENKLVLYQFRGVLHQVWV 540
Qy 541 LDSEALRIPKQSHRIDTDG 559
Db 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 3

```
US-10-408-765A-941
/ Sequence 941, Application US/10408765A
/ Publication No. US20040101874A1
/ GENERAL INFORMATION:
/ APPLICANT: Ghosh, Soumitra S.
/ APPLICANT: Fahy, Bojin D.
/ APPLICANT: Zhang, Bing
/ APPLICANT: Gibson, Bradford W.
/ APPLICANT: Taylor, Steven W.
/ APPLICANT: Glenn, Gary M.
/ APPLICANT: Warnock, Dale E.
/ TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
/ FILE REFERENCE: 660088.465
/ CURRENT APPLICATION NUMBER: US/10/408,765A
/ NUMBER OF SEQ ID NOS: 3077
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 941
/ LENGTH: 559
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-408-765A-941
```

```
Query Match 99.7%; Score 2848; DB 16; Length 559;
Best Local Similarity 99.6%; Pred. No. 2.8e-262;
Matches 557; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 MATALSEEBELNDYYSILNVRREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Db 1 MATALSEEBELNDYYSILNVRREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Qy 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREFFERLOREBERRLQ 120
Db 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREFFERLOREBERRLQ 120
Qy 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIENKMHISQSIEAPLTATDTAIL 180
Db 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIENKMHISQSIEAPLTATDTAIL 180
Qy 181 SGLSTONGNGGSSINPALRRVTSAGKMGELFAGDLOGPLFGKLPRLTFRCEVTNN 240
Db 181 SGLSTONGNGGSSINPALRRVTSAGKMGELFAGDLOGPLFGKLPRLTFRCEVTNN 240
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQIG 300
```

Qy 301 IPHSFALISYOHKFQDDQTRVKSGSLKAGFGTVVEGAERKISRHSVLGAASVGPBG 360  
Db 301 IPHSFALISYOHKFQDDQTRVKSGSLKAGFGTVVEGAERKISRHSVLGAASVGPBG 360  
Qy 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATGPIVVPAMRLIIKPYLRAQKEKEL 420  
Db 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATGPIVVPAMRLIIKPYLRAQKEKEL 420  
Qy 421 EKQRESAATDVLQKKQBAESAARLMQSVRRRIIEAESRMGLIIVNAMYGFVNDKSRKS 480  
Db 421 EKQRESAATDVLQKKQBAESAARLMQSVRRRIIEAESRMGLIIVNAMYGFVNDKSRKS 480  
Qy 481 EKVAVIVTVPLQCLVNDKSKILLTEASKAGLPGFYDPCVGEKMLKVLXQFGLVHQMV 540  
Db 481 EKVAVIVTVPLQCLVNDKSKILLTEASKAGLPGFYDPCVGEKMLKVLXQFGLVHQMV 540  
Qy 541 LDSEALRIPKQSHRIDTDG 559  
Db 541 LDSEALRIPKQSHRIDTDG 559

RESULT 4  
US-10-000-897-69  
; Sequence 69, Application US/10000897  
; Publication No. US20030165852A1  
; GENERAL INFORMATION:  
; APPLICANT: Schueler, P.  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE  
; FILE REFERENCE: 11012-004-999  
; CURRENT APPLICATION NUMBER: US/10/000,897  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: 60/248,862  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 69  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-000-897-69

Query Match 60.3%; Score 1723; DB 14; Length 336;  
Best Local Similarity 99.1%; Pred. No. 2.6e-155;  
Matches 333; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKHRDPELKSQAERLFLN 60  
Db 1 MATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKHRDPELKSQAERLFLN 60  
Qy 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTPAETREPERLQREBERRLQ 120  
Db 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTPAETREPERLQREBERRLQ 120  
Qy 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQIEINKMHSQSIETAPLATDAIL 180  
Db 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQIEINKMHSQSIETAPLATDAIL 180  
Qy 181 SGLSLTONGNGGGSINFALRRVTSKAGMGELEFGAGLOGPLFGKLFRMLTPRCFTYN 240  
Db 181 SGLSLTONGNGGGSINFALRRVTSKAGMGELEFGAGLOGPLFGKLFRMLTPRCFTYN 240  
Qy 241 CALQFSRGRIRPGITVTLARNLDKNTVGYLQMRWGIOSANNTSIVRDTKTSHTVALQLG 300  
Db 241 CALQFSRGRIRPGITVTLARNLDKNTVGYLQMRWGIOSANNTSIVRDTKTSHTVALQLG 300  
Qy 301 IPHSFALISYOHKFQDDQTRVKSGSLKAGFGTVVE 336  
Db 301 IPHSFALISYOHKFQDDQTRVKSGSLKAGFGTVVE 336

RESULT 5

US-09-764-868-780  
; Sequence 780, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 780  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (107)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (143)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (144)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (152)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (155)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-780

Query Match 26.3%; Score 750; DB 9; Length 159;  
Best Local Similarity 92.5%; Pred. No. 5.1e-63;  
Matches 147; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Qy 2 ATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKHRDPELKSQAERLFLN 61  
Db 2 ATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKHRDPELKSQAERLFLN 61  
Qy 62 VHOAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTPAETREPERLQREBERRLQ 121  
Db 62 VHOAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTPAETREPERLQREBERRLQ 121  
Qy 122 RTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQIEIN 160  
Db 122 RTNPKGTISVGVDATDLFDRYDXKLKMCRAVXKPEXIEIN 159

RESULT 6  
US-09-764-868-1154  
; Sequence 1154, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1154  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (31)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE

LOCATION: (58)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (105)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (107)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (143)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (144)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (152)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (155)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1154

Query Match 25.7%; Score 733; DB 9; Length 159;  
Best Local Similarity 90.6%; Pred. No. 2,2e-61;  
Matches 144; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 2 ATALSEELDNEDYSLNVRREASSELKAAVRLCMYHPDKRDPKLSQARLFWL 61  
DB 1 ATALSEELDNEDYSLNVRREASSELKAAVRLCMYHPDKRDPKLSQARLFWL 60  
QY 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 121  
DB 61 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 120  
QY 122 RTRPKGTISGVNATDLPFRYDEYEDVSSSPQIEIN 160  
DB 121 RTRPKGTISGVNATDLPFRYDXKXLMCPAVXFPXIEIN 159

## RESULT 7

US-10-437-963-131723  
Sequence 131723, Application US/10437963  
Publication No. US2004012343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 131723  
LENGTH: 569  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_33762C.1.pep  
US-10-437-963-131723

Query Match 23.1%; Score 661; DB 16; Length 569;  
Best Local Similarity 30.2%; Pred. No. 1,2e-53;  
Matches 178; Conservative 103; Mismatches 242; Indels 66; Gaps 14;

QY 8 EELDNDYISLVNREASSELKAAVRLCMYHPDKRDPKLSQARLFWLHQAYE 67  
DB 6 EELDNDYISLVNREASSELKAAVRLCMYHPDKRDPKLSQARLFWLHQAYE 65

QY 68 VLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 127  
DB 66 VLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 124  
QY 128 TISGVNATDLPFRYDEYEDVSSSPQIEINKHISQISAPPLATDAILSGSLSTQ 187  
DB 125 TISGVNATDLPFRYDEYEDVSSSPQIEINKHISQISAPPLATDAILSGSLSTQ 170  
QY 188 NNGGGSINPALARITSAKMGELFPGADLOGPLFGKLFNLPKRVTTNCLQPS 247  
DB 171 GTDGTGAASAVLRHOLSSV--ASVEFMATGRLSISVQTFRQISPHSTATGLSLSD 228  
QY 248 RGIRPGLATVLARNLDKNTVGYLQWRMGIOSAMNTSIYDTSISFTVALQGIPIHSPAL 307  
DB 229 GST--NLSNMTROSLDNI VGNIGLALGTDSSISGMQKDEKSNAGDVKLTGTFGAS 286  
QY 308 ISYQHKFODDQTRVKSILKAGFPGTVEYGAERKISRHSVLGAASVGVV----- 358  
DB 287 AHYTRVFTSKSHGRVAG--RVGSTALDFEIGGGRISERSVTMIVNIGIQISILELDT 344  
QY 359 -----QVSLKVLNRSQTYFPPIHLT---DQLPSAMFYATVGP 396  
DB 345 VINKVGMIMESIVIGVEGVSMPFELHRAQKLVIPVLTSTDPNALLATSVFAL---P 401  
QY 397 LVVYFAMRLIKPY-LRAQKEKEKORESATDVLOKQOASAVRLMOESVRITEA 455  
DB 402 STLYFLQYFPKPYCLKREKQELKFM-ESLSQUTERRAKAOKILFVSNRKNR 460  
QY 456 EESRMGLIIVNAMYG--KRVNDSKRSKSE-----KVKVIDVTVPLQCLYKDS-KLILTEA- 506  
DB 461 QLEDGLVITKALYGNRKVKESSESNELNDVSAQVLDVTLPLNPLVSEAGQLKLHGI 520  
QY 507 SKAGLPGFYPCVGEERKULKYQFRGVLYHQVMVLDSEALRIPKQSHRI 555  
DB 521 KSGIMGFYPCVGEERKULKYQFRGVLYHQVMVLDSEALRIPKQSHRI 569

## RESULT 8

US-09-927-738-17  
Sequence 17, Application US/09927738  
Patent No. US20020076799A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongwen  
TITLE OF INVENTION: Compositions and Methods of modulating TGF- $\beta$  signaling  
FILE REFERENCE: 17633/1082  
CURRENT APPLICATION NUMBER: US/09/927,738  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: US 60/119786  
PRIOR FILING DATE: 1999-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/03561  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent version 3.1  
SEQ ID NO 17  
LENGTH: 118  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Clone S3 + 125 Protein Sequence  
US-09-927-738-17

Query Match 16.0%; Score 456.5; DB 9; Length 118;  
Best Local Similarity 83.5%; Pred. No. 3,2e-35;  
Matches 91; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 451 RTIEAESRMGLIIVNAMYGFVNDKSRKSEKVIYDVTPLOQCVKDSKLITREASRAG 510  
DB 11 RTIEAESRMGLIIVNAMYGFVNDKSRKSEKVIYDVTPLOQCVKDSKLITREASRAG 69  
QY 511 LFGFYDPCVGEERKULKYQFRGVLYHQVMVLDSEALRIPKQSHRIDTDG 559  
DB 70 LFGFYDPCVGEERKULKYQFRGVLYHQVMVLDSEALRIPKQSHRIDTDG 118

```
RESULT 9
US-10-000-897-71
; Sequence 71, Application US/10000897
; Publication No. US20030165852A1
; GENERAL INFORMATION:
; APPLICANT: Schueler, P.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE
; FILE REFERENCE: 11012-004-999
; CURRENT APPLICATION NUMBER: US/10/000,897
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,882
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 71
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-897-71

Query Match      15.8%; Score 451; DB 14; Length 103;
Best Local Similarity 89.3%; Pred. No. 8,66-35;
Matches 92; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 79 DIYKRGLEMGWEVERRRTPAIRIEPERLQREERERLQORTNPKGTISGVGDATDL 138
D 1 DSVSKKKKKKHVRRRRTPAIRIEPERLQREERERLQORTNPKGTISGVGDATDL 60
QY 139 FDRYDEEYDVSSGSPFOIEINKMHSOSTEAPLTADTTL 181
D 61 FDRYDEEYDVSSGSPFOIEINKMHSOSTEAPLTADTTL 103

RESULT 10
US-10-424-599-225707
; Sequence 225707, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 225707
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(338)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_45845C.1.pep
US-10-424-599-225707

Query Match      12.7%; Score 364; DB 12; Length 338;
Best Local Similarity 27.7%; Pred. No. 1,1e-25;
Matches 106; Conservative 54; Mismatches 146; Indels 76; Gaps 11;

QY 8 EEIDNEHYSLWNRASSEELKAAYRLCMLYHPKMDPELKSAEPLFNLVHOAYE 67
D 6 EAQDNRLVALNLSPASDSBEIRRAYQWAQYHPKYOAPHKDIATENFORICAYE 65
QY 68 VLSDPQTRAIYDIYKRGLEMGWEVERRRTPAIRIEPERLQREERERLQORTNPKG 127
D 68 VLSDPQTRAIYDIYKRGLEMGWEVERRRTPAIRIEPERLQREERERLQORTNPKG 127
```

```
Db 66 ILSDPNKRQIYDIYKRGLE-TSGLELGPKNLNGABEIKABLERLKMKEKREKMAAHQPSG 124
QY 128 TISGVGDATDLFDRYDEEYDVSSGSPFOIEINKMHSOSTEAPLTADTTL 187
D 125 TIVANMSL-----PRYLIDNGM-----LGMAMTSEIQQLSKRNAVTTIGNLAVN 170
QY 188 NGNGGGSINFLARVTSAKMGELFPGAGDLQPLFGKLFRLLTRCFPTTNCALQFSS 247
D 171 GEEGGGATVAFPHHLSVSSVAVASAG--LRALIGVQTRNLSSHAATMGIALSKD 228
QY 248 RGIRPGLTIVLANLKDNTVGLQWNRWGIQSAMNTSIVBDTKTSHFTALQGIPIHSFAL 307
D 229 GSL--NLSNMTQSL-----ETASGHIELALE-----PHSSVT 260
QY 308 ISYQHKFQDDQTRVKSGLK--AGFGTYVEY-----GAB 340
D 261 VGMQCK--DORTSASGEVFGTGSFETSVHYTHRSPKSLGCIIVRVGSSYLEVEVG 317
QY 341 RKISRHSVLGAANVGVQGVGS 362
D 318 KKLKSFYSVWMLYITGI-XGIS 338

RESULT 11
US-10-424-599-184546
; Sequence 184546, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184546
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137661C.1.pep
US-10-424-599-184546

Query Match      11.0%; Score 314; DB 12; Length 234;
Best Local Similarity 39.0%; Pred. No. 3,7e-21;
Matches 90; Conservative 33; Mismatches 96; Indels 12; Gaps 7;

QY 335 VEYGAERKISRHSVLGAANVGVQGVSLKVKLNRASTQYFPPIHLTDOLPSAMFYATV 394
D 6 VEVGGGRKLSKFSFVWMLYITGI-QGISWFEIYRGQKLIPIILTRHLNIPVATGAFV 64
QY 395 GELVYFAMHRLIKR-YLRQKEKELEKREESADVDYQKQGEASAVRLMESVRRI 453
D 65 VPASLYFVRLKQIFIKRYLRRNQKALB-EKERTSAQVREARYAQAQNLQONVANRRK 123
QY 454 EAEESRMGLITIVAWYG--KFVNDKSRKSE-----KRVYIDVVPLOCTLVKDS-KLILTE 505
D 124 NKQLEHFGVIMALYGNQRIINNLKSSSETSELTSEVYIDVTIPLNPLVNDSGQLKHE 183
QY 506 -ASKAGLPGFYDPCVGEERKLVLYQFRGVLYQWVLDSBALRIPKQSHRI 555
D 184 GVKKSGIMGCDPYPSPSPDLVYEVYAGNQYRWAGDYBEIQIPQGSRI 234

RESULT 12
US-10-425-114-71459
; Sequence 71459, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
```





QY 165 SQSIEAPLTATDTAILS-GSISTONGNG 192  
DB 150 --SFGSGSPDFTGTSTFSL---GHGG 172

## RESULT 15

US-10-369-493-47  
; Sequence 47, Application US/10369493  
; Publication No. US2003033675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 47  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Aquifex aeolicus  
US-10-369-493-47

## Query Match

7.8%; Score 222.5; DB 15; Length 376;

Best Local Similarity 26.4%; Pred. No. 4.2e-12; Matches 102; Conservative 44; Mismatches 120; Indels 121; Gaps 19;

QY 13 EDVYSLNVRREASSEELKAAYRRLCMYHPDKRDEPKSQAEKLFNLVHQAYEVLSDP 72  
DB 7 KDVEIILGVPRNNAQOEIKKAYRLVRKHYDICKPE---CEKKEINEAYQVLSDP 62  
QY 73 QTRAIYDIYKRGJLMBG-WEVERRRTP--AEIRBEF-----ERLQREERERLQORT 123  
DB 63 EKRLDYMYGHAAPGAGAOQREVETTEIPLEIRREFDFDGSIFERATGRRRRRR 122  
QY 124 NPKG-TISGVADATDLFDRVDEEYEDVSGSFPQIEINKMHISQIEAPLTATDTAILS 182  
DB 123 SVKSEDIVPEIT---LEAFK---GTVP-IEVER-----EVPSCAC----- 159  
QY 183 SLSTONGNGGSSINFALRVTSAKGGELFEAGDLOGLFGLKLFNLTFRCPVTNCA 242  
DB 160 -----GCTGYDSKSRCTCTCGRGE-----TVQGNWF--FQVRQTCPTC----- 197  
QY 243 LQSSSRGIRPGLTTLARNLDKNT-VGYLQWRMGIQSAMNTSIYDRTKSHFTVALQGI 301  
DB 198 -----GGEGVIIYENCHACTGRGY-----GLVKEI-----IKVKI 226  
QY 302 PHSFALISYQHKFPDDDOTRKYKSLKAGFFG-----TVVEYGAERKISR----- 345  
DB 227 PPG-----VVDGSKLVVEGKGHAGRGPGPDLYIIVKVKPKHIFERKGDLYVDV 277  
QY 346 -----HSLVGAASVGVQGVSLKVKL 367  
DB 278 NITPEAVLGTVEVPTLDGKVKVKI 304

Search completed: July 13, 2004, 08:52:42  
Job time : 354 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 27 Seconds  
(without alignments)  
1991.521 Million cell updates/sec

Title: US-09-787-678A-6  
Perfect score: 2857  
Sequence: 1 MATALSEEDLNEDYSLN.....VLDSALRIPKSHRIDPTDG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646	22.6	591	2 T29974	hypothetical prote
2	607.5	21.3	539	2 S58287	J-domain protein D
3	606	21.2	537	2 B84772	probable DnaJ prot
4	437.5	15.3	642	2 T41504	hypothetical DnaJ
5	239.5	8.4	369	2 B72327	dnaJ protein - The
6	235.5	8.2	328	2 G71610	protein with DnaJ
7	224.5	7.9	349	2 B86237	protein F14N23.23
8	222.5	7.8	376	2 E70361	chaperone DnaJ - A
9	218	7.6	354	2 F71623	protein with DnaJ
10	211	7.4	337	2 G84590	probable heat shoc
11	205.5	7.2	350	2 T06689	heat shock protein
12	205	7.2	242	2 JC7933	spermatogenic cell
13	204	7.1	348	2 T04618	heat shock protein
14	201	7.0	335	2 T48161	heat shock protein
15	195	6.8	392	2 B86495	heat shock protein
16	195	6.8	392	2 G72128	heat shock protein
17	194	6.8	340	2 JN0912	heat-shock protein
18	194	6.8	376	2 AH211	DnaJ protein [impo
19	192.5	6.7	331	2 A96621	probable heat shoc
20	192.5	6.7	418	2 S42031	LBD2 protein - lee
21	192	6.7	178	2 G02272	heat shock protein
22	192	6.7	420	2 T49127	dnaJ protein homol
23	192	6.7	420	2 S71199	heat shock protein
24	190	6.7	389	2 A64202	heat shock protein
25	189	6.6	390	2 S73459	heat shock protein
26	188.5	6.6	423	2 T09358	dnaJ-like protein
27	187.5	6.6	419	2 D81242	DnaJ protein homol
28	187.5	6.6	419	2 T07371	dnaJ protein homol
29	187.5	6.6	499	2 G96831	hypothetical prote

30	187	6.5	351	2 S23508	dnaJ protein homol
31	185.5	6.5	355	2 T40385	hypothetical prote
32	185.5	6.5	910	1 S73361	dnaJ homolog prote
33	185	6.5	377	2 A83052	DnaJ protein PA476
34	184.5	6.5	369	2 D64686	co-chaperone and h
35	184.5	6.5	391	2 G84611	probable DnaJ prot
36	183.5	6.4	346	2 T48049	hypothetical prote
37	182	6.4	336	2 T09133	heat shock protein
38	182	6.4	373	2 F81333	chaperone DnaJ Cj1
39	181	6.3	417	2 UQ2142	chaperone DnaJ pro
40	180.5	6.3	332	2 S76622	hypothetical prote
41	180.5	6.3	376	2 G90630	DnaJ protein [impo
42	180.5	6.3	376	2 G85481	chaperone with Dna
43	180.5	6.3	392	2 H71526	probable heat shoc
44	180.5	6.3	392	2 D81693	dnaJ protein homol
45	180.5	6.3	419	2 T01643	DnaJ protein homol

ALIGNMENTS

RESULT 1	
T29974	hypothetical protein F11G11.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans	
C:Date: 15-Oct-1999	#sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29974	
R:Latreille, P.; Deadman, R.	
A:Submitted to the EMBL Data Library, November 1996	
A:Description: The sequence of C. elegans cosmid F11G11.	
A:Reference number: Z20715	
A:Accession: T29974	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-591 <LAT>	
A:Cross-references: EMBL:U80451; PIDN:AA37835.1; GSPDB:GN00020; CESP:F11G11.7	
A:Experimental source: strain Bristol N2; clone F11G11	
C:Genetics:	
A:Gene: CESP:F11G11.7	
A:Map position: 2	
A:Introns: 36/3; 213/3; 288/1; 541/3; 585/1	
Query Match	22.6%; Score 646; DB 2; Length 591;
Best local Similarity	30.2%; Pred. No. 6.7e-38;
Matches 181; Conservative 110; Mismatches 231; Indels 78; Gaps 19;	
QY	7 EEBLDNEDYSLNVEDREASSEEIKAYRRLCMVHPDKRDELSQAERLFTNLVHOAY 66
DB	22 EEBI---DFAILNVPRKATDDDEIKAYRRCIMFHDRVNDDEKKDAERVFVKLRRAH 78
QY	67 EVLSDPQTRAIYDIYKRGKLEMEGWEYVERRRTPAIREEFERLQREERERLQORTNPK 126
DB	79 EVLIDPKORAIYVALGQGLDTQGMELVSANPENIRKEVEFLQRIKRELMLORHPT 138
QY	127 GTISGVDAATDLPDRYDEBYEDVSGSFPQIEINKHISQSIETAPLTAFTALISGSIST 186
DB	139 SAFWIKTTIGMFQENDE-----RYPP-QLLGISLSGSVDCATGVDPRFGLSERVKT 190
QY	187 QNGNGGSSINFALRYVTSAGKMGELERFAGDLQGPL-----FGIKLFRNLTPRCFVT 238
DB	191 GNGKGDGVSAGVAKVVA-----GSVNLNNTISLSAESVSLTCRAARVFTTRAIVI 240
QY	239 TNCALQPS--SRGIRGLTTLVLANIDK-----NTVGY--LQWR-- 274
DB	241 VOPQLQYNMLHEAIIPIISIAMSMNLRIKKVYSEKRLFTDSIPCKIPFYSMRLLHTRQGS 300
QY	275 -----GIOSAMTSTYRDKTSHFTVALQL---GIHPSPALSYQHKKFDDQTRVXGS 325
DB	301 IVNLISPIANALTTTWH--TENNHAAYVGLTSPINSVRLVYTRKENDSITTEW--S 357
QY	326 LKAGFGTVEYGAERKIRSHSVLGAASVGVPOG--VSLKVKLNRAISQTYFPPIHLLTD-- 382
DB	358 VQLTYYGVNPAISMRLSRISRSISCSFHSFSCLLYTKFKLAKAGSTFDWQIVLDCDK 417



C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T41504  
R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: 221999  
A:Accession: T41504  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-642 <MOL>  
A:Cross-references: EMBL:AL049522; FIDN:CAB40007.1; GSPDB:GN00068; SPDB:SPCC63.03  
C:Genetics:  
A:Experimental source: strain 972h-; cosmid c63  
A:Gene: SPDB:SPCC63.03  
A:Map position: 3  
A:introns: 18/3; 423/2; 563/3

Query Match 15.3%; Score 437.5; DB 2; Length 642;  
Best Local Similarity 24.0%; Pred. No. 4.5e-23;  
Matches 151; Conservative 120; Mismatches 246; Indels 111; Gaps 20;

QY 8 EELNEDYSLNLRRASSSEELKAAYRLCLYHPDKRDELKSAERLFNLVHQAYE 67  
DB 2 DEADSEWELYLALGPKDATSDQIKESYRLRLPHPDH-TADQKAABEKFOIIGAYE 60  
QY 68 VLSDPQTALIDYIKRGLMEGWYVERRRTPAEIRREFFRLQREBERRLQORTNPKG 127  
DB 61 VLSDPKKEIYDNFGEQGLKTD-NWVGFPKGSABELKKKIREQIOERDIHRIDSLVQSR 119  
QY 128 TISVGVDTDFDRDEVEDVSG-----SSPQIENKMHISQIEAPLATD 176  
DB 120 ETTIVNMTPLFAR-NIVQNALGLGAGTRMLTPYERFSLIQWVSFOIKSSFSIPTSPSN 178  
QY 177 -----TALISGLSTQ-----NGNGGSINFALRYTSAKMGEL----- 211  
DB 179 DLKRFSPFSFGSGDFDSAPSDDEBNHTSSSLSLVTEASMRQNSKLOPSIFAYVHS 238  
QY 212 -----EFGAGDIQGLPLFGKLFRNLTPRCFVTTNCALOPSSRGIRPGLITVLAARNLD 263  
DB 239 QPSPNLSSSEIGFSLRPLGLITVKSVAIINNQTFIVP-LIQISGLKRPQATVIGRQIT 296  
QY 264 KNTGVYLQMR-----WGIQSANNTSIVADTKTSHTVALQ--LGIPIHS----- 304  
DB 297 RFGITLAAWKKGWVSLSGWGIASPRGAN-----SSFSLSLQOMKALPNSLVPQLSMNA 349  
QY 305 -----FALISYQHFKQ---DDQTRVK-GLSKAGFGTVVEYGAERKISRHSVLGAAY 353  
DB 350 EVTGLAMVSGLAIVNNLKNATEDSPYQIKLGTSMTVGGLQVSGDTSKVKGRYITFGVNI 409  
QY 354 SVGVAPQG-VSLKVKLNRASTQYFFPIHLTDQLPSAMFYATGVPLVYFAMRLLIKP-- 410  
DB 410 SVGVAPGTGITSILNMSRSLGQKISLPIWMCVSFSDASVFWGVLPITTSILGVEQFLRPR 469  
QY 411 -----YLAQKEKELEKQRESAATDVLOKKQAEBAVYLMQESVRRIIEAESRKL 462  
DB 470 LSNQRLRLRLQKLDKDSQ-----RRKQSAIRAVVLMKEIYEVKKQKLEMEKGL 519  
QY 463 IIVNAMYKFPNDSKRSKSEKVIDVTPVLCVLVDSKL-ILTEASAKAGLPGFYDPCVGE 521  
DB 520 VIEYAEY-RVNVCGANEDPKQ--DVTISIALVENSRLAIPSSVKSISITGIPLPSDN 576  
QY 522 EKNLKVLYQFRGVLIHQVWVLDSEALRIPKQ 551  
DB 577 EKLEIYVTFHQQRHRRVVLRLDKQGVFLPSR 606

RESULT 5  
B72327  
dnaJ proteain - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72327  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.C.M.

Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between *Plasmodium falciparum* and *Bacteria* from genome sequence

A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: B72327

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-369 <ARN>

A:Cross-references: A:AE001751; GB:AE000512; NID:g4981371; PIDN:AA035931.1; PID:g4981371

A:Experimental source: strain MSB8

C:GeneticSB:

A:Gene: TM0849

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:7-73/Domain: dnaJ amino-terminal homology <DNU>

Query Match 8.4%; Score 239.5; DB 2; Length 369;

Best Local Similarity 22.5%; Pred. No.2.2e-09;

Matches 103; Conservative 62; Mismatches 155; Indels 139; Gaps 17;

QY 11 DNEITYSLNVRREASSEELKAAYRRLCMLYHPDKHRDPKLSQAEURLFNLVHQAYEVL 70

DB 4 EKDYYEILGVPDRAIOEIRKAYKRLVKNHMPDRH--PENRKEAEQRFKEIQEAYEVL 61

QY 71 DPQRAIYDIYKKGLEMEGHEVVERRRTPAEIKKEEPRLL-----QRRERR 118

DB 62 DPQRRAMYDRFGYVG-EQPTVQETESGGFPDDIFDFENIRNDIFDVFFGGRPHQERR 120

QY 119 IQQRTNPFGTISVGADATDLPDRYDEYEDVSGSFPOIEINKHISQSIAPLATDTA 178

DB 121 EYARGEIDIRKIEIYTLSDL-----INGAEIP-VERYRY----- 153

QY 179 ILSGLSTQNGNGGGSINPALRVTSAKWELEFGAGDLQGLFGLKLFNLTGRCVYT 238

DB 154 ---ETCRGCGTGVAPNAGYMDCPSCGGRIR---EERRSPFGYFVSERTCERCGGT 205

QY 239 TNCALQF-----SRGIRPGLTIVLANLDKATVGYIQMRKGIOSAMNTSYVRPTKISH 292

DB 206 GKIPREYCHECGSGGRVLRK-----VRRT----- 229

QY 293 FTVALQGIHPSPALISYQHKFODDQTRVGLSKAGPFG-----TVVEXGARKKISR 345

DB 230 -----VKIPRVVEDGTHLRITGGGNAGYGGPYGDLITIVRYKPPRRFK 274

QY 346 -----HSVLGAASVGVPOG--VSLKVKLNRASTQTFEPIHLTDQLPSAM 389

DB 275 SGGSLDVYDVTIDYQALIGTTVEVPLPEGGTMTLKIPIGTQPEYF---RLKGKGLPVR- 330

QY 390 FYATVGPLVVFYFAMHRLTIKPYLPAQKE--KELEKOR 424

DB 331 -YGRRGDLIV--NVHVEIPKSLSRERKRVLEELAKR 364

RESULT 6

protein with DnaJ domain, DNU1/SIS1 family PFB0595w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: G71610

R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Patter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

A>Title: Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: G71610

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-328 <GAR>

A:Cross-references: GB:AE001405; GB:AE001362; NID:g3845224; PIDN:AA071908.1; PID:g3845224

A:Experimental source: clone 3D7

C:GeneticSB:

A:Gene: PFB0595w

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F/4-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 8.2%; Score 235.5; DB 2; Length 328;  
Best Local Similarity 24.9%; Pred. No. 3.6e-09;  
Matches 64; Conservative 38; Mismatches 64; Indels 91; Gaps 7;  
13 EDVYSLNVRREASSELKAAVRLCMLYHPDKHDPDLKSOAERLFNLVHQAYEVLSDP 72  
3 KDYVSLIGVSDCTNDTKKAYRKLMMHPDKHDEKSKCKEAEKFKNAEAYDLADE 62  
73 QTRAIYDIYGRKGLMEGWVERRRRTPAERIEEFERLQREERERLQORTNPKGTISV 131  
63 EKRIYDTYGEELG-----KGSIPFG 83  
132 -----GVNATPLFDR-----YDEYEDVS-----GSSFPQIEIN 160  
84 GNTYVYSGVPSSELPFSRIFSSDQFSTSTFDEDFSPSTFVMTSRKSRPSTTNTIN 143  
161 KMHISQSIAPLTATDTAILSGS-----LSTONGGGSINFALRYTSAGWG 209  
144 NKKPRTAYEVPPLSLSEELSGCKKKLIRKRFRMGKSYEDNNTYIDVK-----AGMK 198  
210 E----LEFGAGDIQGPL 222  
199 DGTKITFYVGEEDQLSPM 215

## RESULT 7

E86237  
protein F14N23.23 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: E86237  
R/Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chun, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nansen, N.F.; Hughes, B.; Hultz, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Schatz, G.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: E86237  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-349 <STO>  
A/Cross-references: GB:AE005172; NID:g4914317; PIDN:AAD32885.1; GSPDB:GN00141  
C/Genetics:  
A/Gene: F14N23.23  
A/Map position: 1  
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 7.9%; Score 224.5; DB 2; Length 349;  
Best Local Similarity 31.1%; Pred. No. 2.4e-08;  
Matches 60; Conservative 22; Mismatches 62; Indels 49; Gaps 5;  
14 DYVSLNVRREASSELKAAVRLCMLYHPDKHDPDLKSOAERLFNLVHQAYEVLSDP 73  
4 DYVNVLKVNANEDDKSKYRRMAKWHDPK--NPLSKKAEAKFKFOISEADVLSDP 61  
74 TRAIYDIYGRKGLMEGWVERRRRTPAERIEEFERLQREERERLQORTNPKGTISV 133  
62 RRQIYDYGEELKSTD-----LPTAETLAHQOQRSSSSNEFRYPR----- 106  
134 DATDLFDRYDEYEDVSGSSFPQIEINMHISQSIAPLTATDTAILSGSLSTONGGGS 193  
107 DAEDFAEFEGEGDAFGG-----SSGRTREDGDGGG 140  
194 SINPALRYTSAG 206  
141 -----RRFKSAE 147

## RESULT 8

E70361  
chaperone DnaJ - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 20-Aug-1999  
C/Accession: E70361  
R/Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: E70361  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-376 <AQF>  
A/Cross-references: GB:AE00703; NID:g2983287; PIDN:AAC06881.1; PID:g2983289; GB:AE00065  
C/Genetics:  
A/Gene: dnaJ2  
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F/8-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 7.8%; Score 222.5; DB 2; Length 376;  
Best Local Similarity 26.4%; Pred. No. 3.6e-08;  
Matches 102; Conservative 44; Mismatches 120; Indels 121; Gaps 19;

13 EDVYSLNVRREASSELKAAVRLCMLYHPDKHDPDLKSOAERLFNLVHQAYEVLSDP 72  
7 KDYVSLIGVPRNNSQBEIKKAYRKLVRKYHPDICKPE-----CEEKKEINBAVQLSDP 62  
73 QTRAIYDIYGRKGLMEGWVERRRRT- AEIEEF-----ERLQREERERLQORT 123  
63 EKRIYDTYGEELGHAFFBAGAQOQVETTELPIRELLREFDFDGSIFERATGRDARRR 122  
124 NPKG-TISGVNATPLFDRYDEYEDVSGSSFPQIEINMHISQSIAPLTATDTAILSG 182  
123 SVKGEIVVPVIT-----LEAFK---CTYTP-IEVER-----EYVCSAC----- 159  
183 SLSTONGGGSINFALRYTSAGWGELEFGAGDQGLFELKLFRLTTPRCFVTTNCA 242  
160 -----GGTGYDESKSRFCPCGGRGE-----TVQGNWF--FQVRQCPYC----- 197  
243 LQSSKGRIPGLTYLAARLDKNT--VGYQMRWGSQSMNTSIVDDTKSHHTVALQGL 301  
198 -----GEGVAYENCACTGRGY-----GLVRET-----IKVKI 226  
302 PHSFALISYQHKFODDQTRVKSGLKAGFFG-----TVVEYGAERKISR----- 345  
227 PPG-----VRGSKLVVBGKGAHAGRGPPDDVYIIVAKRHKIFERKGDLLYDV 277  
346 -----HSVIGAAVSVGVPOGVSLLKYL 367  
278 NITYPEAVLVGTEVEVPTLDGEKVKYKI 304

## RESULT 9

F71623  
protein with DnaJ domain PF80090c - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C/Accession: F71623  
R/Gardner, M.J.; Tectelin, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Petter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; MUID:99021743; PMID:9804551  
A/Accession: F71623  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-354 <GAR>  
A/Cross-references: GB:AE001370; GB:AE001362; NID:g3845089; PIDN:AAC71808.1; PID:g384509

A: Experimental source: clone 3D7  
C: Genes: PP80090c  
A: Gene: PP80090c  
C: Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F: 21-89/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 7.6%; Score 218; DB 2; Length 354;  
Best Local Similarity 27.7%; Pred. No. 6, 9e-08;  
Matches 56; Conservative 29; Mismatches 49; Indels 68; Gaps 4;

OY 14 DYVSLNVRREASSEELKAAVRLCMLYHPDKGRDPPELKSQAERLFNLVHQAIVEVLSDPQ 73  
|||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 21 DYVSLGVSDCTNEDDKKAKKLAMKHPDKHLNAASKGADNNFKSISEAYEVLSDSE 80  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 74 TRAIYDIYGRKGLEMEGWEEVERRRTPAEIRIEEFERLQREERRRIQORTNPKTGISVG 133  
|||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 81 KRDIYDYKYGEGGLDKYG-----SNHGSKGFRTDP----- 111  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 134 DATDLDRYDEBEVDVGSSRFPQIEINGMHSQSIEAPLTATDTLLSGSLSTQNQN--- 190  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

DB 112 -NDVSKFKPK-----TEYFRYSPPSPNGNVLP 139  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 191 ----GGGINPALRVTSAGKM 208  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

DB 140 EGSLFGGSSPFPSGINPRSGSGY 161  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 10  
G84590  
probable heat shock protein [imported] - Arabidopsis thaliana  
C: Species: Arabidopsis thaliana (mouse-ear cress)  
C: Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C: Accession: G84590  
R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
Neus, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter,  
Nature 402, 761-768, 1999  
A: Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A: Reference number: A84420; PMID:20083487; PMID:10617197  
A: Accession: G84590  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residue: 1-337 <STO>  
A: Cross-references: GB:AE002093; MID:94586038; PIDN:AAD25656.1; GSPDB:GN00139  
C: Genes:  
A: Gene: At2g20560  
A: Map position: 2  
C: Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 7.4%; Score 211; DB 2; Length 337;  
Best Local Similarity 54.1%; Pred. No. 2e-07;  
Matches 40; Conservative 13; Mismatches 19; Indels 2; Gaps 1;

OY 14 DYVSLNVRREASSEELKAAVRLCMLYHPDKGRDPPELKSQAERLFNLVHQAIVEVLSDPQ 73  
|||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 4 DYVKVLQVDSASDDDKAKRKLAMKWHPDK--NPNNKKDAEAMFKQISSEAYEVLSDPQ 61  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 74 TRAIYDIYGRKGLE 87  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

DB 62 KKAIVDYQYGEGLK 75  
|||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 11  
T06689  
heat shock protein homolog T17F15.190 - Arabidopsis thaliana  
C: Species: Arabidopsis thaliana (mouse-ear cress)  
C: Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jan-2000  
C: Accession: T06689  
R: Querrier, F.; Choiane, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattoilico, L.; Artiguen  
submitted to the Protein Sequence Database, April 1999  
A: Reference number: Z15793  
A: Accession: T06689  
A: Molecule type: DNA

A:Residues: 1-350 <Q9E>  
A:Cross-references: EMBL:AL049658  
A:Experimental source: Cultivar Columbia; BAC clone T17F15  
C:Genetics:  
A:Map position: 3  
A:Introns: 203/2  
A:Note: T17F15.190  
C:Superfamily: heat shock protein dnaJ, dnaJ amino-terminal homology  
F:4-71/Domain: dnaJ amino-terminal homology <DND>

Query Match 7.2%; Score 205.5; DB 2; Length 350;  
Best local similarity 26.5%; Pred. No. 5,2e-07;  
Matches 62; Conservative 27; Mismatches 72; Indels 73; Gaps 7;

Dy LSVLLVNRREASSEELKAA YRRLCMLYHPDKRDPELKSGAERLFNLYHQAYEVLSDPQ 73  
| | | | |  
Db DYNIILKNHNAETEDDLKKA YKRLLAMIMHPDKNPSTR-RDEAEAKFKRISEAVDVLSDPQ 62  
| | | | |  
74 TRAIYDIYGRKGL-----MEGVEVERRRTPAIRIEEFRLQREBEERLQORTPKG 127  
| | | | |  
63 KRQIYDYGEERGKLKSGKI PNSSSSEASSSSSSRYPFHQRPHPPNVASSFENPR- 121  
| | | | |  
128 TISGVVATDFDRYDEYEVDGSSSPQEIINKMHISQSI EAPLTATDTALLSGLSIQ 167  
| | | | |  
122 -----DAEDT-----YAFPSSE----- 134

Oy 188 NGNGGSINFALRRVTSAKCGIELF-----GAGDLGGPLFGTLFNNLP 233  
| | | | |  
Db 135 --NGGGSNN-----AGGRGNRAFRNGHFNTGANGSYGEMRKVPAMENP 178

RESULT 12  
JC7933  
spermatogenic cell-specific dnaJ-like protein, MFSU1 protein - Japanese macaque  
C:Species: Macaca fuscata (Japanese macaque)  
C:Date: 22-Jun-2003 #sequence\_revision 22-Jun-2003 #text\_change 07-Jul-2003  
C:Accession: JC7933  
R.Yu, S.S.; Takenaka, O.  
Biochem. Biophys. Res. Commun. 301, 443-449, 2003  
A>Title: Molecular cloning, structure, and testis-specific expression of MFSU1, a member  
A:Reference number: JC7933; MUID:22452769; PMID:1256581  
A:Accession: JC7933  
A:Molecule type: mRNA  
A:Residues: 1-242 <YUA>  
A:Cross-references: DDBJ:AB095737  
C:Comment: This protein is a testis-specific dnaJ-like protein. It acts together with tr  
atogenesis.  
C:Genetics:  
A:Gene: mfsj1  
C:Keywords: DnaJ-like protein; spermatogenesis; testis-specific

Query Match 7.2%; Score 205; DB 2; Length 242;  
Best local similarity 32.5%; Pred. No. 3.4e-07;  
Matches 64; Conservative 28; Mismatches 59; Indels 46; Gaps 8;

Dy LSVLLVNRREASSEELKAA YRRLCMLYHPDKRDPELKSGAERLFNLYHQAYEVLSDPQ 73  
| | | | |  
Db 3 NYIEVLGVQVRPEDIKKAYRKALKMHPDK--NPDNKEARRRPQVAEATEVLSDAK 60  
| | | | |  
74 TRAIYDIYGRKGL-----MEGVEVERRRTPAIRIEEFRLQREBEERLQORE-----E 116  
| | | | |  
61 KRDYDHYGEAGAGSCAVGRPFEDPEYIFSRPDPAEYVFREFGGDDPSDFFGNPLE 120  
| | | | |  
Oy 117 RRLQRTNPKGITISVGDAIDFLRYDEYEVEDVSGSFPIEINKHISISIAPLTATD 176  
| | | | |  
Db 121 NIIASRNRSKSRRG--SAPLFSTFE-----FP-----AFGGFSSFD 158  
| | | | |  
Oy 177 TAIS-GSLSTONGNG 192  
| | | | |  
Db 159 TGFSFSGSL-----GSGG 171

RESULT 13





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 75 Seconds

(without alignments)  
2351.662 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857

Sequence: 1 MATLSEEDLNDDYSLN.....VLDSFAIRPKSHRIDTG 559

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	2848	99.7	559	4	Q9NVH1
2	2826	98.9	552	4	Q8NDM2
3	2792	97.7	559	11	Q8CGU5
4	2552	89.3	507	4	Q86CJ7
5	2461	86.1	563	13	Q802X9
6	1677	58.7	327	11	Q8C1Z4
7	1452	50.8	330	13	Q7SZY3
8	1220	42.7	318	4	Q9UMJ8
9	1160.5	40.6	545	5	Q9V6Z8
10	1022	35.8	214	11	Q8BP83
11	661	23.1	569	10	Q8LN96
12	646	22.6	591	5	P91243
13	607.5	21.3	539	10	Q39079
14	606	21.2	537	10	Q9SL89
15	437.5	15.3	642	3	Q9Y7T0
16	363	12.7	82	4	Q7Z6Y9

17	311	10.9	79	13	Q7SZY4	Q7SZY4 brachydanio
18	235.5	8.2	328	5	Q96212	Q96212 plasmodium
19	229	8.0	401	5	Q96123	Q96123 plasmodium
20	226.5	7.9	349	10	Q8LAU1	Q8LAU1 arabidopsis
21	224.5	7.9	349	10	Q9SY77	Q9SY77 arabidopsis
22	224	7.8	242	6	Q8WN90	Q8WN90 bos taurus
23	220.5	7.7	342	10	Q84SD3	Q84SD3 oryza sativ
24	219.5	7.7	145	4	Q8WRF6	Q8WRF6 homo sapien
25	217	7.6	312	11	Q9DN41	Q9DN41 mus musculu
26	212	7.4	433	5	Q81489	Q81489 arabidodiu
27	211	7.4	337	10	Q9SIL3	Q9SIL3 arabidopsis
28	207	7.2	261	11	Q9DAN5	Q9DAN5 mus musculu
29	205.5	7.2	350	10	Q9SUS7	Q9SUS7 arabidopsi
30	205	7.2	242	6	Q86224	Q86224 macaca fusc
31	204	7.1	348	10	Q49457	Q49457 arabidopsis
32	203	7.1	323	10	Q8L8S5	Q8L8S5 arabidopsi
33	203	7.1	323	10	Q9SR91	Q9SR91 arabidopsi
34	203	7.1	515	10	Q8H812	Q8H812 oryza sativ
35	201	7.0	335	10	Q9M034	Q9M034 arabidopsi
36	200.5	7.0	237	13	Q7ZM83	Q7ZM83 brachydanio
37	196	6.9	273	5	Q7YTD1	Q7YTD1 cryptospori
38	195.5	6.8	424	5	Q8IL88	Q8IL88 plasmodium
39	195	6.8	385	16	Q7VG06	Q7VG06 helicobacte
40	194	6.8	342	10	Q9PEW7	Q9PEW7 lycopersico
41	194	6.8	376	16	Q8YU45	Q8YU45 anabaena sp
42	193.5	6.8	259	5	Q9V7K6	Q9V7K6 drosophila
43	193	6.8	375	16	Q8DKR7	Q8DKR7 synchococc
44	193	6.8	388	16	Q8BUM4	Q8BUM4 mycoplasma
45	192.5	6.7	331	10	Q9XIF5	Q9XIF5 arabidopsi

#### ALIGNMENTS

#### RESULT 1

Q9NVH1 PRELIMINARY; PRT; 559 AA.

AC Q9NVH1; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein FLJ10737.  
OS Homo sapiens (Human).  
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isegal T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Negahari K., Masuno Y.,  
RA Minomura K., Iwanaga T.,  
RT "NDO human cDNA sequencing project";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK001599; BAA91780.1; -;  
DR HSBP; P25685; IHDJ.  
DR InterPro: IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ\_1.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PSS0076; DnaJ\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 559 AA; 63336 MW; A610P997833276A0 CRC64;

Query Match 99.7%; Score 2848; DB 4; Length 559;

Best Local Similarity 99.6%; Pred. No. 2.3e-211;

Matches 557; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
1 MATLSEEDLNDDYSLNVRREASSEELKAYRRLCMVYHPDKRDPETKQARLTF 60  
|||||

```
Db 1 MATALSEELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKHDPDLKSQAERLFLN 60
Qy 61 LVHQAAYVSDPOTRAIYDIYGRKGLEMEGWEVERRRTPAEIREEERLQREERERLQ 120
Db 61 LVHQAAYVSDPOTRAIYDIYGRKGLEMEGWEVERRRTPAEIREEERLQREERERLQ 120
Qy 121 QRTNPKGTISVGDATDLFDRYDEEYEDVSGSSFPQIEINKMHSQSIEAPLATATALL 180
Db 121 QRTNPKGTISVGDATDLFDRYDEEYEDVSGSSFPQIEINKMHSQSIEAPLATATALL 180
Qy 181 SGSLSTONGGGSINPALRRVTSAKWGELFEGAGDLOGPLFGLKLFNNLTPRCVTTN 240
Db 181 SGSLSTONGGGSINPALRRVTSAKWGELFEGAGDLOGPLFGLKLFNNLTPRCVTTN 240
Qy 241 CALQFSSRGIRPGLITVLARNLDKNVGYLQWRMGISQSMNTSIYVDTKTSHTTVALQIG 300
Db 241 CALQFSSRGIRPGLITVLARNLDKNVGYLQWRMGISQSMNTSIYVDTKTSHTTVALQIG 300
Qy 301 IPHSFALISYOHKFQDDQTRVKSJLKAQFPGTVVEYGAEKRSRHSVLGAASVGPQGS 360
Db 301 IPHSFALISYOHKFQDDQTRVKSJLKAQFPGTVVEYGAEKRSRHSVLGAASVGPQGS 360
Qy 361 VSLKVKLNRASTQYFFPIHLTDOLPSAMFYATVGPVLYVYFAMHRLIIPYLAQKEKEL 420
Db 361 VSLKVKLNRASTQYFFPIHLTDOLPSAMFYATVGPVLYVYFAMHRLIIPYLAQKEKEL 420
Qy 421 EKQRESAATDVLOKKQAEASAVRLMOESVRRIIEAESRMGLIIVAMYGKFPVNDKSRKS 480
Db 421 EKQRESAATDVLOKKQAEASAVRLMOESVRRIIEAESRMGLIIVAMYGKFPVNDKSRKS 480
Qy 481 EKVAVIDVTVPLQCLVKDSKLLITTEASKAGLPGFYDPCVGEENKLVLYQFQGVHLQWV 540
Db 481 EKVAVIDVTVPLQCLVKDSKLLITTEASKAGLPGFYDPCVGEENKLVLYQFQGVHLQWV 540
Qy 541 LDSEALRIPEQSHRIDTDG 559
Db 541 LDSEALRIPEQSHRIDTDG 559
```

## RESULT 2

```
Q8NDM2 ID Q8NDM2 PRELIMINARY; PRT; 552 AA.
AC Q8NDM2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP727C181.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast cancer;
RA Pucsek A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL833841; CAD38701.1;
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 552 AA; 62574 MW; 085A3748F5285C0C CRC64;
```

```
Query Match 98.9%; Score 2826; DB 4; Length 552;
Best Local Similarity 100.0%; Pred. No. 1,1e-209;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 EELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKHDPDLKSQAERLFLN 67
|||||
```

```
Db 1 EELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKHDPDLKSQAERLFLN 60
Qy 68 VUSDPTRAIYDIYGRKGLEMEGWEVERRRTPAEIREEERLQREERERLQ 127
Db 61 VUSDPTRAIYDIYGRKGLEMEGWEVERRRTPAEIREEERLQREERERLQ 120
Qy 128 TISVGDATDLFDRYDEEYEDVSGSSFPQIEINKMHSQSIEAPLATATALLSGSLSTQ 187
Db 121 TISVGDATDLFDRYDEEYEDVSGSSFPQIEINKMHSQSIEAPLATATALLSGSLSTQ 180
Qy 188 NNGGGSINPALRRVTSAKWGELFEGAGDLOGPLFGLKLFNNLTPRCVTTN 247
Db 181 NNGGGSINPALRRVTSAKWGELFEGAGDLOGPLFGLKLFNNLTPRCVTTN 240
Qy 248 RGRPELITVLARNLDKNVGYLQWRMGISQSMNTSIYVDTKTSHTTVALQIGIPHSFAL 307
Db 241 RGRPELITVLARNLDKNVGYLQWRMGISQSMNTSIYVDTKTSHTTVALQIGIPHSFAL 300
Qy 308 ISYOHKFQDDQTRVKSJLKAQFPGTVVEYGAEKRSRHSVLGAASVGPQGS 367
Db 301 ISYOHKFQDDQTRVKSJLKAQFPGTVVEYGAEKRSRHSVLGAASVGPQGS 360
Qy 368 NRASQTYFFPIHLTDOLPSAMFYATVGPVLYVYFAMHRLIIPYLAQKEKEL 427
Db 361 NRASQTYFFPIHLTDOLPSAMFYATVGPVLYVYFAMHRLIIPYLAQKEKEL 420
Qy 428 AADVLOKKQAEASAVRLMOESVRRIIEAESRMGLIIVAMYGKFPVNDKSRSEK 487
Db 421 AADVLOKKQAEASAVRLMOESVRRIIEAESRMGLIIVAMYGKFPVNDKSRSEK 480
Qy 488 VTVPLQCLVKDSKLLITTEASKAGLPGFYDPCVGEENKLVLYQFQGVHLQWV 547
Db 481 VTVPLQCLVKDSKLLITTEASKAGLPGFYDPCVGEENKLVLYQFQGVHLQWV 540
Qy 548 IPQSHRIDTDG 559
Db 541 IPQSHRIDTDG 552
```

## RESULT 3

```
Q8C6U5 ID Q8C6U5 PRELIMINARY; PRT; 559 AA.
AC Q8C6U5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical DnaJ N-terminal domain containing protein.
GN E030019A03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RC MEDLINE=22354683; PubMed=1246851;
RA The RANOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL: AK051156; BAC35287.1;
DR MGD; MGI:2443386; E030019A03RIK.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 559 AA; 63233 MW; DA6B7590EC7D2561 CRC64;
```

```
Query Match 97.7%; Score 2792; DB 11; Length 559;
Best Local Similarity 96.6%; Pred. No. 4.9e-207;
```

Matches	540; Conservative	14; Mismatches	5; Indels	0; Gaps	0
Qy	1	MATALSEELDNEDYYSLINLRREASSSEELKAAYRLCMLTHPDKRPDELKSQAEIRLN	60		
Db	1	MATALSEELDNEDYYSLINLRREASSSEELKAAYRLCMLTHPDKRPDELKSQAEIRLN	60		
Qy	61	LVHQAAYELSDPQTAIVDIYQKGLBMEGMEVVRARRPAIREEFRLRERERELQ	120		
Db	61	LVHQAAYELSDPQTAIVDIYQKGLBMEGMEVVRARRPAIREEFRLRERERELQ	120		
Qy	121	QRTNPKGTISVGVADATDLEDRYDESEYEDVSGSPFOIEINKMHSQSIEAPLTADTAIL	180		
Db	121	QRTNPKGTISVGVADATDLEDRYDESEYEDVSGSPFOIEINKMHSQSIEAPLTADTAIL	180		
Qy	181	SGSLSTQNGNGGGSINPALRRRTSKAGGELFEFGGDIQGLPLKLFRLNLTPRCFVTNN	240		
Db	181	SGSLSTQNGNGGGSINPALRRRTSKAGGELFEFGGDIQGLPLKLFRLNLTPRCFVTNN	240		
Qy	241	CALQSSSGIRPGLTTLVLAARNLDKRTVGLQWRMGIOGAMNTSIYRDRTKSHFTVALQGL	300		
Db	241	CALQSSSGIRPGLTTLVLAARNLDKRTVGLQWRMGIOGAMNTSIYRDRTKSHFTVALQGL	300		
Qy	301	IPHSFALLSYQHKFODDDQDTRVKSGLKAGFEGTVEYGAERKISHSVLGAASVGVQGL	360		
Db	301	IPHSFALLSYQHKFODDDQDTRVKSGLKAGFEGTVEYGAERKISHSVLGAASVGVQGL	360		
Qy	361	VSLLKKNLRAQTYFFPIHLTDQGLPSAMFATVTPVLVYVPMRHLIKPILRAQKEKL	420		
Db	361	VSLLKKNLRAQTYFFPIHLTDQGLPSAMFATVTPVLVYVPMRHLIKPILRAQKEKL	420		
Qy	421	EKQREASATDVLOKQQAESAVALMQESVVRITIEAESRMGLIYNAYGKGFVNDKSRKS	480		
Db	421	EKQREASATDVLOKQQAESAVALMQESVVRITIEAESRMGLIYNAYGKGFVNDKSRKS	480		
Qy	481	EKKVYIDVTVPLOCLYKDSKILTEASVAGLPGFYDPCVGEKMLKVLVYQFQVLAQYV	540		
Db	481	EKKVYIDVTVPLOCLYKDSKILTEASVAGLPGFYDPCVGEKMLKVLVYQFQVLAQYV	540		
Qy	541	LDSEALRIIPKOSHRITDGL	559		
Db	541	LDSEALRIIPKOSHRITDGL	559		
RESULT 4					
Q96CL7 PRELIMINARY: PRT: 507 AA.					
AC	Q96CL7	01-DEC-2001 (TREMBlrel. 19, Created)			
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)				
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)				
DE	Hypothetical protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RL	Strausberg R.;				
RL	Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC014145; AAH14145.1; -				
DR	InterPro; IPR001623; DnaJ_N.				
DR	Pfam; PF00226; DnaJ_1.				
DR	SMART; SM00271; DnaJ_1.				
DR	PROSITE; PS00636; DnaJ_1; 1.				
DR	PROSITE; PS50076; DnaJ_2; 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 507 AA; 57196 MW; 72BD808620B5512A CRC64;				
Query Match 89.3%; Score 2552; DB 4; Length 507;					
Best Local Similarity 90.7%; Pred. No. 1.5e-188;					
Matches 507; Conservative 0; Mismatches 0; Indels 52; Gaps 1;					

Oy	I	MATLSEBELDNEYSLANRRASSEEKAAARRLCMLYHPDKHDPBELKSQAERLFN	60
Dd	1	MATLSEBELDNEYSLANRRASSEEKAAARRLCMLYHPDKHDPBELKSQAERLFN	60
Oy	61	LVAQAYEVLSPPORAIYDIYGKGGLMEGGEVERERTPAIEEFEEFLOREBERRLQ	120
Dd	61	LVAQAYEVLSPPORAIYDIYGKGLMEGGEVVERRTPAIEEFEEFLOREBERRLQ	120
Oy	121	QRINPKGTISVGVDATDLFDRIYEYDVSGSSPPOIEINKMHISQSIEAPLTADTAII	180
Dd	121	QRINPKGTISVGVDATDLFDRIYEYDVSGSSPPOIEINKMHISQSIEAPLTADTAII	180
Oy	181	SGSISTONGNGCGGSINFALNRVTYSAKWGMBLEFCAGPLQGPFGLFRNLTPRCFYTTN	240
Dd	181	SGSISTONGNGCGGSINFALNRVTYSAKWGMBLEFCAGPLQGPFGLFRNLTPRCFYTTN	240
Oy	241	CALOFSSRGIRPGTLTYLARLDKNNTGYLOMRGISANTSTSIVRDTKTSHPTVALQLG	300
Dd	241	CALOFSSRGIRPGTLTYLARLDKNNTGYLOMRGISANTSTSIVRDTKTSHPTVALQLG	300
Oy	301	IPHSFPALISYOXKEODDDQTRVKSKLAGFFGVVEYGAEKRKISRHSVLAQAASVGPQG	360
Dd	301	IPHSFPALISYOXKEODDDQTRVKSKLAGFFGVVEYGAEKRKISRHSVLAQAASVGPQG	360
Oy	361	VSLKVKLNRASQTTFPHILTDOLLBSAMFIATVGPLVVIYFAHRLLIKPYLRQKEKEL	420
Dd	361	VSLKV-----KEL	368
Oy	421	EKOESASATDYLOKKOEAESAVRIMOESSVRITTEAESRMGLIIVNAMYGKFNVDKSRKS	480
Dd	369	EKOESASATDYLOKKOESESAVRMOSSVRTITEAESRMGLIIVNAMYGKFNVDKSRKS	428
Oy	481	EKVVIDVTVPLQCLVNDISKILLTEASKAGLPGFYDCVGBEKNKLVLYOPFGYLHOVMV	540
Dd	429	EKVVIDVTVPLQCLVNDISKILLTEASKAGLPGFYDCVGBEKNKLVLYOPFGYLHOVMV	488
Oy	541	LDSEALARIPKOSHRIIDTGG	559
Dd	489	LDSEALARIPKOSHRIIDTGG	507
RESULT 5			
ID	O802X9	PREDIMINARY;	PRT; 563 AA.
AC	O802X9;		
DT	01-JUN-2003 (TREMBUREl_24, Created)		
DT	01-JUN-2003 (TREMBUREl_24, Last sequence update)		
DE	01-OCT-2003 (TREMBUREl_25, Last annotation update)		
DS	Similar to RIKEN cDNA E030019A03 gene.		
OS	Brechydanio rerio (zebrafish) (Danio rerio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		
CC	Actionopterygii; Neopterygii; Teleostei; Osteariophysi; Cypriniformes;		
NCBI_TaxId=7955;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strasbourg R.;		
RL	Submitted (Feb.-2003) to the EMBL/Genbank/DDBJ databases.		
DR	EMBL, BC046887, AAA46887.1; -.		
DR	InterPro; IPR001623; DnaJ_N.		
DR	Pfam; PF00226; DnaJ_1.		
DR	SMART; SMO0271; DnaJ_1.		
DR	PROSITE; PS00636; DNAJ_1; 1.		
DR	PROSITE; PSS0076; DNAJ_2; 1.		
SO	SEQUENCE 563 AA; 63422 MW; B9EC276CCEB320F CRC64;		
Query Match			
Best Local Similarity 86.1%; Score 2461; DB 13; Length 563;			
Matches 462; Conservative 59; Mismatches 37; Indels 4; Gaps 1;			
Oy	1	MATLSEBELDNEYSLANRRASSEEKAAARRLCMLYHPDKHDPBELKSQAERLFN	60
Dd	1	MAALEEDIEDINDYSELNVRRREAOTEBLKASYRRICMLYHPDKHDPBELKSQAERLFN	60



Db 301 VOLGIPHTFIMMSYQKFODDQTKIXGSV 330

RESULT 8

Q9UM08 PRELIMINARY; PRT; 318 AA.

AC Q9UM08; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Rhodes S., Huckle E.;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL109978; CAB53376.1; -

DR HSSP; P25685; 1HDJ.

DR InterPro; IPR001623; DnaJ\_N.

DR Pfam; PF00226; DnaJ\_1.

DR SMART; SM00271; DnaJ\_1.

DR PROSITE; PS00636; DnaJ\_1; 1.

DR PROSITE; PS50076; DnaJ\_2; 1.

KM Hypothetical protein.

SC SEQUENCE 318 AA; 35610 MW; FT287BD843EBB94 CRC64;

Query Match 42.7%; Score 1220; DB 4; Length 318;

Best Local Similarity 100.0%; Pred. No. 6,6e-66;

Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 MLVHPDGRDELKSGAERLNLVHQAEEVLSDPQTRAIYDIYKRGLEMEGWEVERRR 98

Db 1 MLVHPDGRDELKSGAERLNLVHQAEEVLSDPQTRAIYDIYKRGLEMEGWEVERRR 60

Qy 99 TPARIREFEFLQREERRRQORTNPKTISVGVADTDFDRDEEVEDVSSSPFOIE 158

Db 61 TPARIREFEFLQREERRRQORTNPKTISVGVADTDFDRDEEVEDVSSSPFOIE 120

Qy 159 INKNIHQSIAPLTATDTALISGSLSTONGGGSINFALRYTSKAGMGELEFGAGDL 218

Db 121 INKNIHQSIAPLTATDTALISGSLSTONGGGSINFALRYTSKAGMGELEFGAGDL 180

Qy 219 QGPIFGKLFRLNLTPRCFVTTCALQFSSRGIRPGLTTVARNIDKNTVGLQW 272

Db 181 QGPIFGKLFRLNLTPRCFVTTCALQFSSRGIRPGLTTVARNIDKNTVGLQW 234

RESULT 9

Q9V6Z8 PRELIMINARY; PRT; 545 AA.

AC Q9V6Z8; 08T054;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE CG5531 protein (LUD27406p).

OS CG5531.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydriidae; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Baas A., Bakendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,

RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., May A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Gloddek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iegwam C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kenson J.A., Ketchum K.A.,

RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Waasman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Iegwam C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Pounanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochorko S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutnak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Wengell C.U., Lewis S.E.,

RT "Annotation of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkley;

RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Paclebe J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO03815; AAF58269.2; -;  
 DR EMBL: AY069547; AAJ39692.1; -;  
 DR HSSP: P25685; 1HDJ.  
 DR FlyBase: FBgn0033918; CG8531.  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR Pfam: PF002226; DnaJ; 1.  
 DR SMART: SM00271; DnaJ; 1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 SQ SEQUENCE 545 AA; 60912 MW; F605485B094CDD06 CRC64;

Query Match 40.6%; Score 1160.5; DB 5; Length 545;  
 Best Local Similarity 43.9%; Pred. No. 66-81; Indels 13; Gaps 6;  
 Matches 239; Conservative 116; Mismatches 176;

QY 6 SEBELNEDYSLNVRREASSEELKAAYRLCMLYHPDKRDELKQAEELFNLVHQ 65  
 8 SDAELO-ENYVTLNLPDRTAQINTAYRKQSRMHDPDKLDPDKMAELMFNTKKA 66  
 DB 66 YEVLSDPQRAIYDYGKGLMEGWVERRRTPAIRERERLQREERERLQQRTP 125  
 67 YEVLSDPQRAIYDVGKGLRTEGWEILHRTTPDEIREYERLAQAALERRLOQRTP 126  
 QY 126 KGTISVGDATDLDPRDEEYEDVSGSPQIEINQHSQSLEAPLTDATLISGSIS 185  
 127 KGTITVNTETEPAPDD-----SEMPHEIGSMISLQSLTEAPITKDMIMSGNLY 179  
 DB 186 TONGNGGSGINFALRVTSAGWGELEFGADLQGPLFKLPRLNTPRCFTVTTNCA 245  
 180 SSGNGSGCGFVIGRRILN-KGWEILCAGAGN--GFLGKGGRTLSQKTLTGNTLN 236  
 QY 246 SSGIRGLTTLVARNIDKRTVGYLQWRMGISANNTSYRDTKTSHTFVALOLGPH 305  
 237 RDGCVIPALFSTLAVOLDKHTMGSLLTNAGSOSMSFOLDHSKETYSLSSSLVIGTPHY 296  
 QY 306 ALISYGHKFDODOTVRKGSILKAGFCTVVEYGERKISHSVLGAIVSPOGVSILKY 365  
 297 FGLSYTRKMM-ENELKIKLAKVGTGPMGEYGEKKVSIVTATVSGVSVILKF 355  
 DB 366 KLRASQYFPIHITDOLPSAMFYATVGPLVYVPMHRLIKPYLRQKEKELEKOR 425  
 356 KILRSNGSYFPIHLSDEIVPAVAVSVPIVAMFPIKTVMDPEAEKNIIEVERTK 415  
 QY 426 SAATDVLQKQESAVRLMOESVRRITIEESRMGLIYNANYGKFAVNDKSKSEK 485  
 416 QNEORISAKHEASAAVHLMQATYNNRIMTEELARNGLIIVRAVYGGCTLAG-GRQFKPEOS 474  
 DB 486 IDVTVPGLCVKSKILITLTSKAGLPGFPCVGEKKNLVYORFVGLHGVVLDSEA 545  
 475 LDTVTAIQCVKNGKTLQIHDSSKSDLPGFYDPIGDEKILIRIETTYONQEPVHIHIDNEA 534  
 DB 546 LRIP 549  
 535 LRIP 538

RESULT 10  
 Q8BP83 PRELIMINARY; PRT; 214 AA.  
 AC O8BP83;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DE Hypoetical DnaJ N-terminal domain containing protein  
 DE (fragment).  
 GN E030019A03RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Body;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL: AK077544; BAC36856.1; -;  
 DR MGI: 2443386; E030019A03RIK.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 214 AA; 24077 MW; 4D2C1A23D88D2D88 CRC64;

Query Match 35.8%; Score 1022; DB 11; Length 214;  
 Best Local Similarity 93.5%; Pred. No. 7,2e-71;  
 Matches 200; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY 346 HSVLGAASVGPQGVSLKVKLNRSQTYFPIHITDOLPSAMFYATVGPLVYFAMHR 405  
 1 HSVLGAASVGPQGVSLKVKLNRSQTYFPIHITDOLPSAMFYATVGPLVYLAHVR 60  
 DB 406 LIKPYLRQKEKELEKQESATTVLQKQASAVRLMOESVRRITIEESRMGLIIV 465  
 61 LIIRPIYLRQKEKELEKQESATTVLQKQASAVRLMOESVRRITIEESRMGLIIV 120  
 QY 466 NAMYGKFNVDKSRKQKXVIDVTVPLQCLVDSKILITLASKAGLPGFPCVGEK 525  
 121 NAMYGKFNVDKSRKQKXVIDVTVPLQCLVDSKILITLASKAGLPGFPCVGEK 180  
 QY 526 KYLVQFVGLHGVVLDSEALRIIPKQSHRIDTDG 559  
 181 RVLVQFVGLHGVVLDSEALRIIPKQSHRIDTDG 214

RESULT 11  
 Q8LN96 PRELIMINARY; PRT; 569 AA.  
 AC Q8LN96;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DE Putative DnaJ domain containing protein.  
 GN OSUNB0015K05.11.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Scrophophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_Taxid=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buehl C.R., Yuan Q., Ouyang S., Liu J., Ganeberg K., Jones K.M.,  
 RA Overton II L.L., Tsiltrin T., Kim M.M., Bera J.U., Jin S.S.,  
 RA Fadrosch D.W., Tallon L.J., Koo H., Zisemann V., Heisao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utebbach T.T., Feidlyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quakebush J.,  
 RA White O., Salzberg S.L., Fraser C.M.,  
 RT "Oryza sativa chromosome 10 BAC OSUNB0015K05 genomic sequence."  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA The Rice Chromosome 10 Sequencing Consortium;  
 RT "In-depth view of structure, activity, and evolution of rice  
 RT chromosome 10."  
 RL Science 300:1566-1569 (2003).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buehl C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;



Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC090870; AAM95682.1; -  
DR EMBL; AB017110; AAP54538.1; -  
DR Gramene; OBLN96; -  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ\_1.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
SEQUENCE 569 AA; 63322 MW; BSD44574DBE415AB CRC64;

Query Match 23.1%; Score 661; DB 10; Length 569;  
Best Local Similarity 30.2%; Pred. No. 2.6e-42;  
Matches 178; Conservative 103; Mismatches 242; Indels 66; Gaps 14;

QY 8 EELNEDYYSILNVRREASSELKAYRLCMLYHPDKRDEBELSQARLFLNLYHQA 67  
DB 6 EPEGRRELIALHSPDASGERIRRAYQVQIYHPDKYQDPOMKDVATENFORIRDA 65  
QY 68 VLSPDQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREFERLQREERERLQQRTPNK 127  
DB 66 ILSDENKQIYDIYGMELN-SGLELQPKAKPEIKQLELKRKEEEKFLAHARPTG 124  
QY 128 TISVGVATDLPDRYDEEYEDVSGSFPQIEINKMHISQSIAPLPTATDAILSGSLSTQ 187  
DB 125 SIANFVSVOYLDGYG-----IMRGKMSSEVQLPFSKKNVVGNNLVN 170  
QY 188 NGNGGGSINPALRVTSKAGWELFPGADLQGLPGLKLRNLTPRCFTVTTCALPSS 247  
DB 171 GTDTGTGAASAVLRHQLSSV--ASVEFMAITAGLRSLISVQTRQISPHSTATSGALSLRD 228  
QY 248 RGIRPGLTTLVARNLDRKTVGVQLWRMGIOGAMNTSIYDRKTSHTFTALQGLPHSPAL 307  
DB 229 GSI--NLNMTATRLQSDNIYGNIGLQALGDSISVGMKQDEKKSAAQDVKLGNVFGAS 286  
QY 308 ISYHKFQDDQDTRKYSKAGFPGTVVYGAERKISRHSVILGAASVGV----- 358  
DB 287 AHYRFYFSTKSHGRAV--RVGSTALDPEICGGRISFSTVRMIVNIGINISILEDT 344  
QY 359 -----QGVSLKVKLNKASQTYFFPIHLT---DQLSPAMYATVGP 396  
DB 345 VINKVGMKWEISYIGGEVEGVSWRPELHAGOKLIVPVLSTDFNALATSPAFI--P 401  
QY 397 LVVVFAMRLTIKPY-LPAQKEKELEKORESAATDVLOKQOASAVRLMESRRIIEA 455  
DB 402 STIFLLOTFYVVKCYCKREKQKELEKM-ESLSQULTEARPAKQAQCLTEPVSNNKKNR 460  
QY 456 EESRMGLIIVNAMYG--KFVNDKSRKSE-----KVVIVDTVPLQCLVQDS-KLILTEA- 506  
DB 461 QLEDDGLVITKALYGNRKVKESSESNELNDVASQVLDVTIPLNPLVSEAGQLKLEGI 520  
QY 507 SKAQLPGFYDPCVGBEKNLKVLYQFRGVLYHGVNVLDSALRIPKQSHRI 555  
DB 521 KKSGLMGFYDPCPDPKLLIVEYTFHQYKVMVDYALILPQIHQI 569

## RESULT 12

P91243 PRELIMINARY; PRT; 591 AA.  
AC P91243;  
DT 01-MAY-1997 (TREMBlrel. 03; Created)  
DT 01-MAY-1997 (TREMBlrel. 03; Last sequence update)  
DT 01-OCR-2003 (TREMBlrel. 25; Last annotation update)  
DE Hypothetical protein.  
GN F1G11.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodermidae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN 1;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;

RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).

RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA latreille P.; Deadman R.;  
RT "The sequence of C. elegans cosmid F1G11.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; U80451; AAB37835.1; -  
DR PIR; T29974; T29974.  
DR HSSP; P08622; 1XBL.  
DR WormPep; F1G11.7; CE07060.  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ\_1.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
KW Hypothetical protein.  
SEQUENCE 591 AA; 67122 MW; 751FSB552749C68C CRC64;

Query Match 22.6%; Score 646; DB 5; Length 591;  
Best Local Similarity 30.2%; Pred. No. 3.9e-41;  
Matches 181; Conservative 110; Mismatches 231; Indels 78; Gaps 19;

QY 7 EELNEDYYSILNVRREASSELKAYRLCMLYHPDKRDEBELSQARLFLNLYHQA 66  
DB 22 EEEI--DFYAIINVPDQATDDEIRIKYRRCCLMFDRVDNDEKDAERVFVKLRRAH 78  
QY 67 EVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREFERLQREERERLQQRTPNK 126  
DB 79 EVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREFERLQREERERLQQRTPNK 138  
QY 127 GTISVGVATDLPDRYDEEYEDVSGSFPQIEINKMHISQSIAPLPTATDAILSGSLST 186  
DB 139 SAFMIKTTIGMFQENED-----RYPF-QLLGISGSDVCAPFGVDRPFGSGVKY 190  
QY 187 QNGNGGGSINPALRVTSKAGWELFPGADLQGLPGLKLRNLTPRCFTVTTCALPSS 238  
DB 191 GNGRGDSVSAVWKRAV-----GSYNLENTISLSASVSLTCRAARNVFTRAVI 240  
QY 239 TNCALQPS--SRGIRPGLTTLVARNLDRKTVGVQLWRMGIOGAMNTSIYDRKTSHTFTALQGLPHSPAL 274  
DB 241 VQPOLQYNMLHEALIPSIAMSMNLRIKQVSEKRLFTDSIFCKIIPFYSMRILHTRWQGS 300  
QY 275 -----GIOSAMNTSIYDRKTSHTFTALQGLPGLKLRNLTPRCFTVTTCALPSS 325  
DB 301 IVNLISPIANALTTTVH--TENNHAQVAGSLTSPINSVRLVYIIRKENDSITTEM--S 357  
QY 326 LKAGFPGTVVYGAERKISRHSVILGAASVGVQD--VSLKVKLNKASQTYFFPIHLTD-- 382  
DB 358 VQLTGGVGNPAISMDRRLSRYSRISCSFHSFSCLLYTFKFKKAGGSTFDMQIVLCDDK 417  
QY 383 QLLPSAMFYATVGPVLVYFA--MHRLLIKPYLAQKEKELEKORESAATDVLOKQOAS 440  
DB 418 EALSRSVLY--GVALPYFSFOIAKYIFRMWKEKFSMPEDNSREQVD--VAKKEBAAN 472  
QY 441 AVRLMESVARIIEAESRMGLIIVNAMYGKFVNDKSRKSEKY--KVIVDTVPLQCLVQDS 498  
DB 473 IVSLMATARIRGDESKQGVILSASAKGQCDVSGTRAVPLAGEHTIDVTVLQAMVND 532  
QY 499 SKLILTEASAGLPGFYDPCVGBEKNLKVLYQFRGVLYHGVNVLDSALRIPKQSHRI 558  
DB 533 SOLRVYTV--KSQLPFGFYDPCPGBPKNLTVRYIRGGEHSCVADDEMLMLPLRAHRLSSE 591





Db 284 GLG-----ASARYTRKLSKSHGRIVG-RIGSNALIEBIGGGRQISEFTVMRYVGL 336  
Qy 358 POGVSLKYNKRNASQTYFFPIHLTDLLPSAMFYATGVLVYFAMRLIIPK--LPAOK 416  
Db 337 -KGFVMEVLEHRSQKIVPILLSAHLAPVATGAFVPTSLYFLKKFVVKPILLRREK 395  
Qy 417 EKELEKRESAATDVLOKKQAEASAVRLMOESVRIIEAESRGLIIVNAMYG--KEVN 474  
Db 396 OKALENN-EKTWGVGVARARAEAKOQLQVATRKKNROVETDGLIVTALYGDPRAT- 453  
Qy 475 DKSKSEKVK-----VIDVTPLQCLVKDS-KLILTE-ASKAGLPGFYDPCVGEKMLKV 527  
Db 454 --ERRNMGVGLDSGVLDVTVPMNFVYSDSGQLKLHGVYKSGIMGFCDCPCQPKDLYI 511  
Qy 528 LYQFRGVLAHQVWVLDSEALRIIPKQ 551  
Db 512 AYTYHSQPFVIVGVDEELSIPOE 535

RESULT 15  
Qy770 ID Qy770 PRELIMINARY; PRT; 642 AA.  
AC Qy770;  
DT 01-NOV-1999 (Tremblrel, 12, Created)  
DT 01-NOV-1999 (Tremblrel, 12, Last sequence update)  
DT 01-JUN-2003 (Tremblrel, 24, Last annotation update)  
DE Hypothetical DNAJ domain protein.  
GN SPCC63.03.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL049522; CAB40007.1; -.  
DR PIR; T41504; T41504.  
DR HSSP; P25685; 1HDJ.  
DR GeneDB Spombe; SPCC63.03; -.  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ; 1.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS0076; DnaJ\_2; 1.  
DR Hypothetical protein.  
KM  
SQ SEQUENCE 642 AA; 72287 MW; BA2B8BE1CALED88 CRC64;

Query Match 15.3%; Score 437.5; DB 3; Length 642;  
Best Local Similarity 24.0%; Pred. No. 5.8e-25;  
Matches 151; Conservative 120; Mismatches 248; Indels 111; Gaps 20;

Qy 8 BELDNEVYSLNRRREASSELKAAYRLCMLYHPDKHRDPELKSQAEFLNLYHQAYE 67  
Db 2 DEADSWELVALGLPKATSDQIKESYRLSRLEPHDRH-TADQKAAAEKFOIIOHAYE 60  
Qy 68 VLSDPQTRAIYDIYKGLKLEMGWEVERRTTPAIEEERFRLQREERERRLOQRTNPKG 127  
Db 61 VLSDPSKKEIYDNGEGGLKTD-WNVGFPKSAEELKTKRTOEQERDHEIDSLVQSR 119  
Qy 128 TISGVDAATDLFDRIYDEYEDVSG-----SSFPQIEINKMHIQSISIEAPLTATD 176  
Db 120 ETTIVAMTDLFAF-NIRVQNALGLGAGTRMLTYERFSLIQWVSFOIKSSFISPTFSN 178  
Qy 177 -----TAILSGSLSTQ-----NGNGGGSINPALRVTSAKMGEL----- 211  
Db 179 DLKPSFNSPSSGSGFDDFSPASDEDEGNHKTSSRLSIVTEASRQNSKLQPSIFAVYHS 238  
Qy 212 -----EFGAGDLQGPLFGKLKFLRNLTFRCFVTTCALQFSSRGIRPGITVLANLD 263  
Db 239 QPSFNLSSSEIGFSLRLRGLITVKSVAINNQTPIVP--LTQISGLKRPQATVIVIGRQIT 296

Qy 264 KATVGYLQNR-----WGIOSAMNTSIYRDKTSHFTYALQ--LGIPHS----- 304  
Db 297 RFGTLTARMKTVGWSLGSWGIAISPRGAN-----SSFSLTWQOMKATPNSLVLPQLSMNA 349  
Qy 305 -----FALISYQHKFQ---DDQTRVK-GSLKAGFPGVVEYGAERKISRHSVGAAY 353  
Db 350 EYTAGLMTSGIANNVNLKNAITEDSPYQIKLGTSMSITVGLQVSGDTSRKVGRYSTGVNI 409  
Qy 354 SVGVPOG-VSLKYNKRNASQTYFFPIHLTDLLPSAMFYATGVLVYFAMRLIIPK-- 410  
Db 410 SVGVPTGSITFSLMWSRLQKISLPIIMCQSVFDRSAVFMGLVFPITISILGVEQFRLPRR 469  
Qy 411 -----YLRQKEXELEKQRESAATDVLOKKQAEASAVRLMOESVRIIEAESRGL 462  
Db 470 LSNQKELRLRLRQKLDQSE-----RKKVSAIRAVKMKELVEKKQKLEMERGGL 519  
Qy 463 IIVNAMYGKPVNDKSRSEKVKVITDVTPPLQCLVKDSKL-ILTEASKAGLPGFYDPCVGE 521  
Db 520 VIEYAEY-RVYNGGANEPLDKQ--DVTISIALVENSRLAIPBSVSKSITIGIYPLPSDN 576  
Qy 522 EKNLKVLYQFRGVLAHQVWVLDSEALRIIPKQ 551  
Db 577 EKELEIVYTFHQQRHVRVLRDKGVFLPSR 606

Search completed: July 13, 2004, 08:43:38  
Job time : 82 secs

**This page blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:43:46 ; Search time 12299 Seconds

(without alignments)  
11322.967 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213

Sequence: 1 gaaagctgcgaagatgagc.....gaaagcagaaaaaaaaa 3213

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBml: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: gb\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_hcg\_hum: \*  
31: em\_hcg\_inv: \*  
32: em\_hcg\_other: \*  
33: em\_hcg\_mus: \*  
34: em\_hcg\_pln: \*  
35: em\_hcg\_rtd: \*  
36: em\_hcg\_mam: \*  
37: em\_hcg\_vtc: \*  
38: em\_sy: \*  
39: em\_hcgo\_hum: \*  
40: em\_hcgo\_mus: \*  
41: em\_hcgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3191.4	99.3	3230	9	AF306695 Homo sapi
2	3171.6	98.7	3199	6	AK877081 Sequence
3	3171.6	98.7	3199	6	BD156474 Primer fo
4	3171.6	98.7	3199	9	AK001599 Homo sapi
5	3157.4	98.1	3178	9	HSMB805121
6	1947	60.6	2064	9	AK025491
7	1934	60.2	2125	9	BC014145
8	1851.6	57.6	2454	6	AX748271 Sequence
9	1851.6	57.6	2454	9	AK093709 Homo sapi
10	1845.2	57.4	3012	9	AK095386
11	1517.2	47.2	112158	9	HS126A5
12	1510.8	47.0	191481	2	AC108123
13	1397.4	43.5	2052	6	AX834019
14	1397.4	43.5	2052	9	AK096217
15	1236.4	38.3	1258	9	AF052158
16	1230.2	38.3	1258	9	AF052158
17	1006.6	31.3	1014	9	AF306694
18	989.8	30.8	2774	5	BC046897
19	832.2	25.9	1767	6	AK339329
20	746.2	23.2	1357	9	HS126A512
21	684.2	21.3	706	6	AX868425
22	684.2	21.3	779	6	BD221093
23	575.6	17.9	477	9	HUM267D08
24	466.4	14.5	455	6	BD264940
25	455	14.2	455	6	BD264940
26	455	14.2	455	6	AR238343
27	455	14.2	455	6	AX366559
28	445.2	13.9	550	6	AX873477
29	445.2	13.9	550	6	BD153539
30	385	12.0	386	6	BD221537
31	373	11.6	373	6	AX885936
32	373	11.6	373	6	BD025566
33	331.4	10.3	438	11	G37511
34	306	9.5	821	6	AX554944
35	302.6	9.4	455	6	AX330980
36	300	9.3	300	6	BD219127
37	292.2	9.1	2000	3	AY069547
38	271	8.4	355	9	AF306696
39	200.8	6.2	1411	3	AK115029
40	170.4	5.3	201635	10	AL611931
41	166.2	5.2	171	11	G62036
42	162.8	5.1	237743	2	AC129831
43	139.2	4.3	35217	2	AC017131
44	139.2	4.3	165176	3	AC007588
45	139.2	4.3	180425	3	AC008229

#### ALIGNMENTS

RESULT 1  
LOCUS AF306695 3230 bp mRNA linear PRI 01-OCT-2002  
DEFINITION Homo sapiens clone R5' mRNA sequence.  
ACCESSION AF306695  
VERSION AF306695.1 GI:23393219  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (Bases 1 to 3230)  
XU,H., Sha,Y., Foltz,L., Romero,R., Larsen,A., Mahoney,W. and  
Schueler,P.A.  
TITLE Identification and characterization of genes expressed in human

Pred. No. is the number of results predicted by chance to have a

fetal erythroblasts  
Unpublished  
2 (bases 1 to 3230)  
XU, H., Sha, Y., Foltz, L., Romero, R., Larsen, A., Mahoney, W. and  
Schueler, P. A.  
Direct Submission  
Submitted (18-SEP-2000) Chief Technology Office, Roche Diagnostics,  
2929 7th Street, Suite 100, Berkeley, CA 94710, USA  
Location/Qualifiers  
source  
1. 3230  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="R5"  
/cell\_type="erythroblasts"  
/dev\_stage="fetal"

ORIGIN

Query Match 99.3%; Score 3191.4; DB 9; Length 3230;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 3206; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 GAAGGTTGGGAAGTGGCGAGCGCTTGAAGCGAGAGAGCTGGACATGAAGACTATT 60  
DB 7 GAAGGTTGGGAAGTGGCGAGCGCTTGAAGCGAGAGAGCTGGACATGAAGACTATT 66  
QY 61 ACTGTTGCTGAAGCTGGCGAGGGAGCGCTTCTGTAAGAGCTGAAGAGCTGCTACCGGA 120  
DB 67 ACTGTTGCTGAAGCTGGCGAGGGAGCGCTTCTGTAAGAGCTGAAGAGCTGCTACCGGA 126  
QY 121 GGCTGTGATGCTTCAACATCCAGACAGACAGAGAGAGAGCTCAAGTCAAGAGCGG 180  
DB 127 GGCTGTGATGCTTCAACATCCAGACAGAGAGAGAGCTCAAGTCAAGAGCGG 186  
QY 181 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAAGTGTATGACCCCAACAGAGG 240  
DB 187 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAAGTGTATGACCCCAACAGAGG 246  
QY 241 CCATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGGGAAA 300  
DB 247 CCATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGGGAAA 306  
QY 301 GGAAGAGAACCCCTGCTGAATTCAGAGAGATTGAGCGCTGCAGAGAGAGAGAG 360  
DB 307 GGAAGAGAACCCCTGCTGAATTCAGAGAGATTGAGCGCTGCAGAGAGAGAGAG 366  
QY 361 AGAGAGATTCAG 420  
DB 367 AGAGAGATTCAG 426  
QY 421 CCGACCTTTTGAATGCGTATGATGAGAGATGAGAGATGATGTCGGGAGTATGCTTCGG 480  
DB 427 CCGACCTTTTGAATGCGTATGATGAGAGATGAGAGATGATGTCGGGAGTATGCTTCGG 486  
QY 481 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGAGAGAGAGAGAGAGAG 540  
DB 487 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGAGAGAGAGAGAGAGAG 546  
QY 541 ACACAGCATCTCTCTGGAAGCTCTCAACCCAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 547 ACACAGCATCTCTCTGGAAGCTCTCAACCCAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 601 ACTTGGCTCAGACGAGTAACCTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 607 ACTTGGCTCAGACGAGTAACCTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
QY 661 GAGACCTACAGGGGCTTTGTCGCTCAAGCTGTTCCGTAACTCAACCAAGATGCT 720  
DB 667 GAGACCTACAGGGGCTTTGTCGCTCAAGCTGTTCCGTAACTCAACCAAGATGCT 726  
QY 721 TTGTGACAAACAACTGTGCTCTGAGTTTCAATCCGTGGAATCCGAGCTGACCA 780  
DB 727 TTGTGACAAACAACTGTGCTCTGAGTTTCAATCCGTGGAATCCGAGCTGACCA 786

QY 781 CTGTCTAGCTGGAACTAGACAAGAACACCGTGGGCTTACCTGCAAGTGGGAGTGA 840  
DB 787 CTGTCTAGCTGGAACTAGACAAGAACACCGTGGGCTTACCTGCAAGTGGGAGTGA 846  
QY 841 TCCAGTACAGCAATGAACACTAGCATGTCGAGACATMAAACAGAGCACTTCACTGTG 900  
DB 847 TCCAGTACAGCAATGAACACTAGCATGTCGAGACATMAAACAGAGCACTTCACTGTG 906  
QY 901 CCTGAGCTGGAAATCCCTCACTCTTGTGACATGATCAGTATACAGCAAAATCCAG 960  
DB 907 CCTGAGCTGGAAATCCCTCACTCTTGTGACATGATCAGTATACAGCAAAATCCAG 966  
QY 961 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTGGAGAGTGTG 1020  
DB 967 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTGGAGAGTGTG 1026  
QY 1021 AGTACGAGCTGAGAGAAATCTCCAGGACAGCGTTTGGGTGACGCTGCAGCGTTG 1080  
DB 1027 AGTACGAGCTGAGAGAAATCTCCAGGACAGCGTTTGGGTGACGCTGCAGCGTTG 1086  
QY 1081 GAGTTCACAGGGCGTTTCTTCAAAAGTCAAGCTCAACAGGGCAGTCAATACTTCT 1140  
DB 1087 GAGTTCACAGGGCGTTTCTTCAAAAGTCAAGCTCAACAGGGCAGTCAATACTTCT 1146  
QY 1141 TCCCTATTCATTGACGAGACGAGCTTCTGCGAGCGCATGTTCTATGSCACCGTGGG 1200  
DB 1147 TCCCTATTCATTGACGAGACGAGCTTCTGCGAGCGCATGTTCTATGSCACCGTGGG 1206  
QY 1201 CTCTAGTGTCTACTTGTGCTATGCAACCGTCTGATCAATCAACCTCAGGGCTCAGA 1260  
DB 1207 CTCTAGTGTCTACTTGTGCTATGCAACCGTCTGATCAATCAACCTCAGGGCTCAGA 1266  
QY 1261 AAGAGAGAAATTTGAG 1320  
DB 1267 AAGAGAGAAATTTGAG 1326  
QY 1321 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGCAG 1380  
DB 1327 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGCAG 1386  
QY 1381 AAGAGTCCAGAAATGGGCTCATCATGTCATGCTGATGCGGAGAGATTGTCAATGACA 1440  
DB 1387 AAGAGTCCAGAAATGGGCTCATCATGTCATGCTGATGCGGAGAGATTGTCAATGACA 1446  
QY 1441 AAGAGAGAAAGCGAG 1500  
DB 1447 AAGAGAGAAAGCGAG 1506  
QY 1501 TGAAGACTCGAAGCTCATCTCAAGAGAGCTCCAGAGAGCTGAGGCTGCTGCTTTATG 1560  
DB 1507 TGAAGACTCGAAGCTCATCTCAAGAGAGCTCCAGAGAGCTGAGGCTGCTGCTTTATG 1566  
QY 1561 ACCCGTGTGGGGAG 1620  
DB 1567 ACCCGTGTGGGGAG 1626  
QY 1621 ATCAGGTATGCTGTGACATGAGAGCCCTCCGATACCAAGCAGTCCCAAGATG 1680  
DB 1627 ATCAGGTATGCTGTGACATGAGAGCCCTCCGATACCAAGCAGTCCCAAGATG 1686  
QY 1681 ATACAGATGATTAACCTGCAAGAACAGATTTTAAAGGCGCAAAAATCTTTTCT 1740  
DB 1687 ATACAGATGATTAACCTGCAAGAACAGATTTTAAAGGCGCAAAAATCTTTTCT 1746  
QY 1741 GGGAGTCTACAAATTTGAAATGAATAAAACCAAGACATCAGATGTTTATTTATATTA 1800  
DB 1747 GGGAGTCTACAAATTTGAAATGAATAAAACCAAGACATCAGATGTTTATTTATATTA 1806  
QY 1801 TTATTTATGAAGGTGTACATTTATCATTTATGTGAAGGACATGACAGACCCAGCTT 1860  
DB 1807 TTATTTATGAAGGTGTACATTTATCATTTATGTGAAGGACATGACAGACCCAGCTT 1866

QY	1861	TTGAGGGGTGCTGGGGGTGAGACTGAGGGAGCCCACTGGGAAACCAACTGGAAGCTGGCC	1920
Db	1867	TTGAGGGGTGCTGGGGGTGAGACTGAGGGAGCCCACTGGGAAACCAACTGGAAGCTGGCC	1926
QY	1921	CATGGCTCTTTTCCCAAGATCAGTCTCTGGAGGGAAGGGCTCTGGCCCTGACTCCGCTG	1980
Db	1927	CATGGCTCTTTTCCCAAGATCAGTCTCTGGAGGGAAGGGCTCTGGCCCTGACTCCGCTG	1986
QY	1991	TGTCCGAGCAACGCTGCTGACCCGAGGCCCGCCGCTGTAGTCTTGGCTGGGACTGGA	2040
Db	1997	TGTCCCGAGCAACGCTGCTGACCCGAGGCCCGCCGCTGTAGTCTTGGCTGGGACTGGA	2046
QY	2041	GCTGTCTGTGAGGACCCCTGGCTTCAACAAGAGCTGAGGCCATTTCTGCACTCCAGC	2100
Db	2047	GCTGTCTGTGAGGACCCCTGGCTTCAACAAGAGCTGAGGCCATTTCTGCACTCCAGC	2106
QY	2101	TGAACATGGGAAACAACCTGAAAGACAGAGCCCTCCCGGTCAAGGAGCCTCTGCTGTG	2166
Db	2107	TGAACATGGGAAACAACCTGAAAGACAGAGCCCTCCCGGTCAAGGAGCCTCTGCTGTG	2166
QY	2161	CTGGCTTCCCACTGACCACTCTCTCTGCTGAAATTTACTGCTTGAATCTGGAGCAGATT	2220
Db	2167	CTGGCTTCCCACTGACCACTCTCTCTGCTGAAATTTACTGCTTGAATCTGGAGCAGATT	2226
QY	2221	GCGGGTTTATTAACCTGCTTTTATCTGAGAAACAAGGGTTTGGAAATTAAGTCTCTT	2280
Db	2227	GCGGGTTTATTAACCTGCTTTTATCTGAGAAACAAGGGTTTGGAAATTAAGTCTCTT	2286
QY	2281	TTTCCCACTCCCAAGGCGCTCAAGTCAATTCACCGGCCCTCGGCTTGGGACAGGGT	2340
Db	2287	TTTCCCACTCCCAAGGCGCTCAAGTCAATTCACCGGCCCTCGGCTTGGGACAGGGT	2346
QY	2341	AGTGTACTCCCGATCCAGAGGCGCTGAGCCCTGACACAAGGTGGCTTCCCGTATCCCGATGG	2400
Db	2347	AGTGTACTCCCGATCCAGAGGCGCTGAGCCCTGACACAAGGTGGCTTCCCGTATCCCGATGG	2406
QY	2401	GAAAAAGCCCTGACACAAGCGGGCTTGAAGCTGTGGCTGTCTTCAACCGCTGACAC	2460
Db	2407	GAAAAAGCCCTGACACAAGCGGGCTTGAAGCTGTGGCTGTCTTCAACCGCTGACAC	2466
QY	2461	CCAAGTCAAGAGTGAAGTCTGGGGGAAGGGCAAGCTCAAGAGACAGAACAGAGCGCTTGG	2520
Db	2467	CCAAGTCAAGAGTGAAGTCTGGGGGAAGGGCAAGCTCAAGAGACAGAACAGAGCGCTTGG	2526
QY	2521	CAAGACATCAGACACACCAACCCAAAGGCGTGAAGCCCAAGGCGCGGCGGTGTATCCA	2580
Db	2527	CAAGACATCAGACACACCAACCCAAAGGCGTGAAGCCCAAGGCGCGGCGGTGTATCCA	2586
QY	2581	GCAAGTGAACCTGCAAGCTCCCGCTCTGTCAAGTTCAGACGTCCTTCAAGAAACAAGG	2640
Db	2587	GCAAGTGAACCTGCAAGCTCCCGCTCTGTCAAGTTCAGACGTCCTTCAAGAAACAAGG	2646
QY	2641	CCTGTGCTCCGAGGCTTCTTCAAGCCCTTCTCAAGTGCCTTGGAGTGGAAAT	2700
Db	2647	CCTGTGCTCCGAGGCTTCTTCAAGCCCTTCTCAAGTGCCTTGGAGTGGAAAT	2706
QY	2701	GCAAGGAGATGAAGACCCGCTCAAGGCGTGAAGCTCGGCTGCAATTAAGTTACGTAGG	2760
Db	2707	GCAAGGAGATGAAGACCCGCTCAAGGCGTGAAGCTCGGCTGCAATTAAGTTACGTAGG	2766
QY	2761	CCTGTCTCTCGGGGCTTGAAGATGGACCAATCAGTTCTCTGTCTGAACCCCTCGAGACA	2820
Db	2767	CCTGTCTCTCGGGGCTTGAAGATGGACCAATCAGTTCTCTGTCTGAACCCCTCGAGACA	2826
QY	2821	AGCGCGGCAAGTGTGTGGCTTGAAGCAAGTGGCGCGGGGGGCCCAAGTGTGCGCGGCT	2880
Db	2827	AGCGCGGCAAGTGTGTGGCTTGAAGCAAGTGGCGCGGGGGGCCCAAGTGTGCGCGGCT	2886
QY	2881	CCAGGCCCAACCAAGCTGTGTGTGAAGTCAGGCAACCTCCCAAGCACTGGATATCTAGT	2940
Db	2887	CCAGGCCCAACCAAGCTGTGTGTGAAGTCAGGCAACCTCCCAAGCACTGGATATCTAGT	2946
QY	2941	AACGGCTAAGAACCTCTTCTCTGTGGTTTGAAGACAGTTCGGGTTGTCCAATTTCTGA	3000

Db	2947	AAAGGCTAAGACCTCTCTCTGAGTTTGAAGAGAGCTTGGGTTGTCCATTTCTGTA	3006
Qy	3001	ACATTCATCTCCATTTTAAAAAAGTTTCTCGACGGCCCAACGGCCGAGCCGCGT	3060
Db	3007	ACATTCATCTCCATTTT-AAAAAGTTTCTCGACGGCCCAACGGCCGAGCCGCGT	3065
Qy	3061	GAGCGTGTGTTCATGAGCTGGGCCCCGGGCGTCCCGTGCCTCTGCTGCGAGTGCT	3120
Db	3066	GAGCGTGTGTTCATGAGCTGGGCCCCGGGCGTCCCGTGCCTCTGCTGCGAGTGCT	3125
Qy	3121	TCGCGGACCCATCTCTCGCTTCATTTGAGTCGACTGACAGAAAGGCACTCACACA	3180
Db	3126	TCGCGGACCCATCTCTCTCGCTTCATTTGAGTCGACTGACAGAAAGGCACTCACACA	3185
Qy	3181	ATTAACCTTCTCTGAAAGCAAAAAA-3213	
Db	3186	ATTAACCTTCTCTGAAAGCAAAAAA-3218	
RESULT 2			
LOCUS	AX877081	3199 bp, DNA	linear PAT 17-DEC-2003
DEFINITION	Sequence 11986 from Patent EP1074617.		
ACCESSION	AX877081		
VERSION	AX877081.1	GI:40031817	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	1		
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
TITLE	Primers for synthesising full-length cDNA and their use		
JOURNAL	Patent: EP 1074617-A 11986 07-FEB-2001.		
FEATURES	Research Association for Biotechnology (JRP)		
source	location/Qualifiers		
	1..3199		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	16..1695		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAB89562.1"		
	/db_xref="GI:40031818"		
	/translation="MATPLSBEELDNEEDYSLNVRREASSSEELKAYRRLCMLYHPD		
	KHRPBLKSOAERLFLNHOAYEVLSPORAYLDYIGKGLMEGWVVERRTPE		
	IRSEFRLQREERRLRQQRNTKPGTISVGDADLDRYDEDEYVSGSPQIEIN		
	KHNSISLAVPLATDVALISGLSTNGGGSINFLARVTSAKMGELFEGADGL		
	OGPFGKLPRLNLRPCFTVALQGLSSGIRPGLTTLVARNIDKNTMGVLIQMRGCI		
	OSANETIVRDTKTSHTFTVALQLGIPSPALLIYQKRPDDDDQTRVKSILKAGFQIV		
	VEYAEKRISRHSVLGAASVGVPOGSLVKLRASQTFPLHLDLPLSAMFTVA		
	TVGLVYVFWMRLLIRPYRAQEKELKQRESAADIVLQKQLEASAVRLQESVRE		
	RIIAEESRMGLILIVNMGKFEVDSKSEKYEIVTVPLQCLIVDSKLILTEASK		
	AGLPGFPCPGEEKNLKYLQPGVLIQVWLDSALRIPKSHRIDTGG"		
ORIGIN			
Query Match	98.7%; Score 3171.6; DB 6; Length 3199;		
Best Local Similarity	99.7%; Pred. No. 0;		
Matches 3188; Conservative	0; Mismatches 9; Indels 1; Gaps 1;		
Qy	2	AAAGTTCGGAAGTGGGAGCGGCTTGAGCGAGAGAGCTGACAAATGAAGCTATTA	61
Db	3	AAAGTTCGGAAGTGGGAGCGGCTTGAGCGAGAGAGCTGACAAATGAAGCTATTA	62
Qy	62	CTCGTGTGAAGCTGCGAGAGAGGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGAG	121
Db	63	CTCGTGTGAAGCTGCGAGAGAGGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGAG	122
Qy	122	GCTCTGTATGCTTACCAATCCAGACAAAGACAGACCCAGAGCTCAAGTCAAGCGCGGA	181

[illegible]

QY	1262	AGAGAGGAATTTGGAGAAAGCAGAGAGGAAAGCGCCGCCACCGATGTGTGCAAGAAAGCA	1321
Db	1263	AGAGAAGGAATTTGAGAGAGCAGAGAGGAAAGCGCGCACCGATGTGTGCAAGAAAGCA	1322
QY	1332	AGAGGCGGAGTCCGCTGTCCGCTGAGTACAGGAATCTGTCCGAAGATTAATTGAGCGAGA	1381
Db	1333	AGAGGCGGAGTCCGCTGTCCGCTGAGTACAGGAATCTGTCCGAAGATTAATTGAGCGAGA	1382
QY	1382	AGAGTCCAGAAATGGGCTCATATCGTCAATGCTGTGTAAGGGAATTTGTCAATGACAA	1441
Db	1383	AGAGTCCAGAAATGGGCTCATATCGTCAATGCTGTGTAAGGGAATTTGTCAATGACAA	1442
QY	1442	GAGCAGGAAGAGCCAGAAAGGTGAAGTGAATTGAAGTACATGTGCCCTTCGACAGTCTCGT	1501
Db	1443	GAGCAGGAAGAGCCAGAAAGGTGAAGTGAATTGAAGTACATGTGCCCTTCGACAGTCTCGT	1502
QY	1502	GAAGGACTCGAAGCTCATCTCGAGAGGCTCCMAAGCGTGGCTGCTGACTGCTTTATGGA	1561
Db	1503	GAAGGACTCGAAGCTCATCTCGAGAGGCTCCMAAGCGTGGCTGCTGACTGCTTTATGGA	1562
QY	1562	CCCGTGTGTGGGGGAAGAAGAAACCTGAAGTCTATACAGTTCCGGGGCGCTCGCA	1621
Db	1563	CCCGTGTGTGGGGGAAGAAGAAACCTGAAGTCTATACAGTTCCGGGGCGCTCGCA	1622
QY	1622	TCAGGTGATGTGTGTGACAGTGTAGGCGCTCCGGATACCAAGACATCCACAGAGATCGA	1681
Db	1623	TCAGGTGATGTGTGTGACAGTGTAGGCGCTCCGGATACCAAGACATCCACAGAGATCGA	1682
QY	1682	TACAGATGAGATAACTGCGCAAGAACCAAGATTTTAAAAAGCCGCAAAAAATCTTTTCTGT	1741
Db	1683	TACAGATGAGATAACTGCGCAAGAACCAAGATTTTAAAAAGCCGCAAAAAATCTTTTCTGT	1742
QY	1742	GGAGTCAACAATTTGGAAAAATGAAAAAACCCAGATCAAGTGTTTTATTTATATATAT	1801
Db	1743	GGAGTCAACAATTTGGAAAAATGAAAAAACCCAGATCAAGTGTTTTATTTATATATAT	1802
QY	1802	TATTTATGAAGGTGTATCCATTATCAATTATGTGAAGGGAACATGACAGACACCCCACTTT	1861
Db	1803	TATTTATGAAGGTGTATCCATTATCAATTATGTGAAGGGAACATGACAGACACCCCACTTT	1862
QY	1862	TGAAGGGTGTGGGGGTGTGACTGAGGGAGGCCCATGTGGGAACCAAGATGACAGCTGGGCC	1921
Db	1863	TGAAGGGTGTGGGGGTGTGACTGAGGGAGGCCCATGTGGGAACCAAGATGACAGCTGGGCC	1922
QY	1922	ATGGCTGTTTTCCCAAGGATCAGTTCCTGGAAGGGAAGGGCTGTGGCCCTGATCTCCGCTGT	1981
Db	1923	ATGGCTGTTTTCCCAAGGATCAGTTCCTGGAAGGGAAGGGCTGTGGCCCTGATCTCCGCTGT	1982
QY	1982	GTCCCGAGACACGTGTCTGACCCGAGGCCGCGCTGTATGTTCTTTGGCTGGGTCTGGAG	2041
Db	1983	GTCCCGAGACACGTGTCTGACCCGAGGCCGCGCTGTATGTTCTTTGGCTGGGTCTGGAG	2042
QY	2042	GTGTCTGTGAGACACCCTGACCTTACACACAGAGAGCTGAGACCACTTTGTGACGTCCAGCT	2101
Db	2043	GTGTCTGTGAGACACCCTGACCTTACACACAGAGAGCTGAGACCACTTTGTGACGTCCAGCT	2102
QY	2102	GAAACATGGAAGAAACAACCTGAAGAACAGAGAGCGCTCCCGGTCAAGGAGCCTCTGTCTGTG	2161
Db	2103	GAAACATGGAAGAAACAACCTGAAGAACAGAGAGCGCTCCCGGTCAAGGAGCCTCTGTCTGTG	2162
QY	2162	TGGGCTTCCCAATGACACCTCTCTGTGTGAATATTTATCTGTGTGAATCTGGAGCGAGTGTG	2221
Db	2163	TGGGCTTCCCAATGACACCTCTCTGTGTGAATATTTATCTGTGTGAATCTGGAGCGAGTGTG	2222
QY	2222	CGGGTTTATATAAATCTGCTTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCGCTTTT	2281
Db	2223	CGGGTTTATATAAATCTGCTTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCGCTTTT	2282
QY	2282	TTCCCACTCCAGAGCTGTCTCAAGTATTCACACCGGCCCTTCGCTTGGGACAGGGTGA	2341
Db	2283	TTCCCACTCCAGAGCTGTCTCAAGTATTCACACCGGCCCTTCGCTTGGGACAGGGTGA	2342

QY 2342 GTGTAACTCCGATCCAGAGGCTAGCCCTGACACAGTGGCTTCCGCTATCCCGTGGG 2401  
 Db 2343 GTGTAACTCCGATCCAGAGGCTAGCCCTGACACAGTGGCTTCCGCTATCCCGTGGG 2402  
 QY 2402 AAAACGCGCTTCCAGACAGGCGCTTGAAGCTGGTGTCTCCCTCAACCGCTGACACAC 2461  
 Db 2403 AAAACGCGCTTCCAGACAGGCGCTTGAAGCTGGTGTCTCCCTCAACCGCTGACACAC 2462  
 QY 2462 CACCTCCAGAGTGAAGTCTGGGACAGGCGCTGACAGAGACAGAGACAGAGCGCTTGGC 2521  
 Db 2463 CACCTCCAGAGTGAAGTCTGGGACAGGCGCTGACAGAGACAGAGACAGAGCGCTTGGC 2522  
 QY 2522 AAGACATGACAGACACCCCAAGGCGTGAAGCCAGGCGCGGCGCTGTATCCAG 2581  
 Db 2523 AAGACATGACAGACACCCCAAGGCGTGAAGCCAGGCGCGGCGCTGTATCCAG 2582  
 QY 2582 CAGGTGACATGACAGCTCCCGCTCTGACAGTCAAGCTCTGACAGAGACACACAGGCG 2641  
 Db 2583 CAGGTGACATGACAGCTCCCGCTCTGACAGTCAAGCTCTGACAGAGACACACAGGCG 2642  
 QY 2642 CTGTGCTCCGAGCGCTTCTTCAAGCTTCTCTCAAGTCCCACTTGGATGCAAGATG 2701  
 Db 2643 CTGTGCTCCGAGCGCTTCTTCAAGCTTCTCTCAAGTCCCACTTGGATGCAAGATG 2702  
 QY 2702 CAGCGAGCTAGAGACCCCTCCAGCGCTGAGACCTCGGCTGAGTAAAGTAAAGTAAAGG 2761  
 Db 2703 CAGCGAGCTAGAGACCCCTCCAGCGCTGAGACCTCGGCTGAGTAAAGTAAAGTAAAGG 2762  
 QY 2762 CTGTCTCTCGGAGCGCTTGAAGTGAAGCTCAAGTGTCTTGTGTCAGACCCCTCGAGCAA 2821  
 Db 2763 CTGTCTCTCGGAGCGCTTGAAGTGAAGCTCAAGTGTCTTGTGTCAGACCCCTCGAGCAA 2822  
 QY 2822 GCGCCGACAGAGTGTGTGCTGAGACAGTGGCGCGGCGGCGGCGGCGGCGGCGGCTC 2881  
 Db 2823 GCGCCGACAGAGTGTGTGCTGAGACAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 2882  
 QY 2882 CAGCCGACCCAGACCTGTGTGCTGAGAGTCAAGCTCCCGACAGCTGATCTGAGTA 2941  
 Db 2883 CAGCCGACCCAGACCTGTGTGCTGAGAGTCAAGCTCCCGACAGCTGATCTGAGTA 2942  
 QY 2942 ACAGCTAAGAACCTCTCTCTCTGATTTTGAAGAGTTCGAGTGTCTGATTTGATA 3001  
 Db 2943 ACAGCTAAGAACCTCTCTCTCTGATTTTGAAGAGTTCGAGTGTCTGATTTGATA 3002  
 QY 3002 CATTCACTCTCATTTTAAAGAGTTTCTGTAAGCGCCCAAGCGCGCGCGGCGGCTG 3061  
 Db 3003 CATTCACTCTCATTTTAAAGAGTTTCTGTAAGCGCCCAAGCGCGCGCGGCGGCTG 3062  
 QY 3062 AGCGTGTGTGATGAGCGCTGCGGCGCTCCGCTGCGCTCTGCGCGAGGCTT 3121  
 Db 3062 AGCGTGTGTGATGAGCGCTGCGGCGCTCCGCTGCGCTCTGCGCGAGGCTT 3121  
 QY 3122 CTGGGACACCATCTCTGCGCTTCAATTGACGTGACGTACAGAGGCACTCACCA 3181  
 Db 3122 CTGGGACACCATCTCTGCGCTTCAATTGACGTGACGTACAGAGGCACTCACCA 3181  
 QY 3182 TAAACCTTTCTGTAAGC 3199  
 Db 3182 TAAACCTTTCTGTAAGC 3199

RESULT 3  
 LOCUS BDI56474 3199 bp DNA linear PAT 17-JAN-2003  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION BDI56474  
 VERSION BDI56474.1 GI:27862232  
 KEYWORDS JP 2002191363-A/11317.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 3199)

AUTHORS Oka, T., Isogai, T., Nishikawa, T., Hayaishi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Makamatsu, A., Nagai, K. and Otsuki, T.  
 TITLE Primer for synthesizing full-length cDNA and use thereof  
 JOURNAL Patent: JP 2002191363-A 11317 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/11317  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI MAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC  
 10,  
 PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 FEATURES Location/Qualifiers  
 source 1..3199  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Query Match 98.7%; Score 3171.6; DB 6; Length 3199;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 3188; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 2 AAAGTTGGCAAGATGGCGACGCGCTTGAAGCGAGAGAGCTGACAAATGAAGCTATTTA 61  
 Db 3 AAAGTTGGCAAGATGGCGACGCGCTTGAAGCGAGAGAGCTGACAAATGAAGCTATTTA 62  
 QY 62 CTGCTGTGATGCTGTACATCTGAGAGAGGCGCTTCTGTAAGAGCTGAAAGCTGCTACCGGAG 121  
 Db 63 CTGCTGTGATGCTGTACATCTGAGAGAGGCGCTTCTGTAAGAGCTGAAAGCTGCTACCGGAG 122  
 QY 122 GCTCTGTATGCTGTACATCTGAGAGAGGCGCTTCTGTAAGAGCTGAAAGCTGCTACCGGAG 181  
 Db 123 GCTCTGTATGCTGTACATCTGAGAGAGGCGCTTCTGTAAGAGCTGAAAGCTGCTACCGGAG 182  
 QY 182 ACAGCTGTAAACCTTGTACCAAGGCTTATGAAGCTTATGATGCCCAACCAAGGCG 241  
 Db 183 ACAGCTGTAAACCTTGTACCAAGGCTTATGAAGCTTATGATGCCCAACCAAGGCG 242  
 QY 242 CATCTATGATATATATGAGAGAGAGGAGGAGCTGGAATGGAAGATGGAGGTTGTGGAAG 301  
 Db 243 CATCTATGATATATATGAGAGAGAGGAGGAGCTGGAATGGAAGATGGAGGTTGTGGAAG 302  
 QY 302 GAGAGAGACCCCTGCTGAATTCGAGAGAGGTTTGAAGCGGCTGACAGAGAGAGAGAGA 361  
 Db 303 GAGAGAGACCCCTGCTGAATTCGAGAGAGGTTTGAAGCGGCTGACAGAGAGAGAGAGA 362  
 QY 362 GAGAGATTTGACAGAGGAGCAATCCCAAGGAGAGATGAGGTTGAGTATATGCCAC 421  
 Db 363 GAGAGATTTGACAGAGGAGCAATCCCAAGGAGAGATGAGGTTGAGTATATGCCAC 422  
 QY 422 CGACCTTTTGTATGATGATGAGAGATGAGAGATGATGATGATGATGATGATGATGATGATG 481  
 Db 423 CGACCTTTTGTATGATGATGAGAGATGAGAGATGATGATGATGATGATGATGATGATGATG 482  
 QY 482 GATTGAATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541  
 Db 483 GATTGAATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 542  
 QY 542 CACAGCCATCTCTCTGAGAGGCTTGAAGCCGAGATGAGATGAGATGAGAGGTTCAATTAA 601  
 Db 543 CACAGCCATCTCTCTGAGAGGCTTGAAGCCGAGATGAGATGAGATGAGAGGTTCAATTAA 602  
 QY 602 CTTTGGGCTGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 661



Db 603 CTTTGCCTCAGACGAGTAACTTTCGGCAAAAGGATGGGAGAGAGTTGGAAATTTTGGAGCTGG 662  
QY 662 AGACCTACAGAGGGCTTTTGTGGTCTCAGCTGTTCCGTAATCTCAACACAGAGTCTT 721  
Db 663 AGACCTACAGAGGGCTTTTGTGGTCTCAGCTGTTCCGTAATCTCAACACAGAGTCTT 722  
QY 722 TGTGACAAACAAGCTGTGCTGTGAGTTTTCATCCCGTGAATCCGACCCGGCTTGACAC 781  
Db 723 TGTGACAAACAAGCTGTGCTGTGAGTTTTCATCCCGTGAATCCGACCCGGCTTGACAC 782  
QY 782 TGTCTAGCTGTGGAACCTAGACAAAGAACACCTGGGCTTACTGCAAGTGGAGTGGGAT 841  
Db 783 TGTCTAGCTGTGGAACCTAGACAAAGAACACCAATGGGCTTACTGCAAGTGGGAT 842  
QY 842 CCACTGAGGCACTAGAACCTAGACCTGTCCGAGACCTAAACAGGCACTTCACTGGC 901  
Db 843 CCACTGAGGCACTAGAACCTAGACCTGTCCGAGACCTAAACAGGCACTTCACTGGC 902  
QY 902 CCTGCAAGCTGGGAATCCCTCACTCCTTGTGCACTGATCACTATCAGACAAATTTCCAGA 961  
Db 903 CCTGCAAGCTGGGAATCCCTCACTCCTTGTGCACTGATCACTATCAGACAAATTTCCAGA 962  
QY 962 TGAAGATCAGACTGTGTGTAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGGTGTGGA 1021  
Db 963 TGAAGATCAGACTGTGTGTAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGGTGTGGA 1022  
QY 1022 GTTACGAGCTGAGAGGAATCTTCAGGCAACGGTTTGGGTGAGCTGTCAAGCTGG 1081  
Db 1023 GTTACGAGCTGAGAGGAATCTTCAGGCAACGGTTTGGGTGAGCTGTCAAGCTGG 1082  
QY 1082 AGTTTCCACAGAGGGCTTCTCTCAAGTCAAGCTCAACAGGGCCAGTCAAGACATCTCTT 1141  
Db 1083 AGTTTCCACAGAGGGCTTCTCTCAAGTCAAGCTCAACAGGGCCAGTCAAGACATCTCTT 1142  
QY 1142 CCTTATTCATCTTGAAGGACAGCTTCTGCCCAGGCCCCATTTCTATGCAACCGTGGGCC 1201  
Db 1143 CCTTATTCATCTTGAAGGACAGCTTCTGCCCAGGCCCCATTTCTATGCAACCGTGGGCC 1202  
QY 1202 TCTAGTGTCTACTTGTGCACTGACCGCTGTGATCATCAACCACTCTGAGGCTGAGA 1261  
Db 1203 TCTAGTGTCTACTTGTGCACTGACCGCTGTGATCATCAACCACTCTGAGGCTGAGA 1262  
QY 1262 AGAAGAGAAATTGGAAGAGCAGAGGAGAAAGCGCCAGCCAGATGTGTGCAAGAAAGCA 1321  
Db 1263 AGAAGAGAAATTGGAAGAGCAGAGGAGAAAGCGCCAGCCAGATGTGTGCAAGAAAGCA 1322  
QY 1322 AGAGGCGGAGTCCGCTGTCCGGCTGATGCAAGAACTGTCTCGAAGATAATTGAGCAGA 1381  
Db 1323 AGAGGCGGAGTCCGCTGTCCGGCTGATGCAAGAACTGTCTCGAAGATAATTGAGCAGA 1382  
QY 1382 AGAGTCCAGAAATGGGCTCTCATCTGTCATGCTGGTACGGGAAGTTGTCAATGACA 1441  
Db 1383 AGAGTCCAGAAATGGGCTCTCATCTGTCATGCTGGTACGGGAAGTTGTCAATGACA 1442  
QY 1442 GAGCAGGAAGAGCAGAAAGGTGAAGTGAATTGAAGTGAAGTGTGCCCCCTGCAAGTGTGAT 1501  
Db 1443 GAGCAGGAAGAGCAGAAAGGTGAAGTGAATTGAAGTGAAGTGTGCCCCCTGCAAGTGTGAT 1502  
QY 1502 GAAAGACTCGAAGCTTCTCTCAAGAGGCTTCCAAAGGCTGGGCTGTCTGGCTTTTATGA 1561  
Db 1503 GAAAGACTCGAAGCTTCTCTCAAGAGGCTTCCAAAGGCTGGGCTGTCTGGCTTTTATGA 1562  
QY 1562 CCGGTGTGTGGGGGAAGAAGAACTGGAAGTGTCTATCACTTCCGGGGCGTCCGTGA 1621  
Db 1563 CCGGTGTGTGGGGGAAGAAGAACTGGAAGTGTCTATCACTTCCGGGGCGTCCGTGA 1622  
QY 1622 TCAAGTGTGTGTGAGACAGTGAAGGCTTCCGGAATCAAAAGAGTCCCAAGATGCA 1681  
Db 1623 TCAAGTGTGTGTGAGACAGTGAAGGCTTCCGGAATCAAAAGAGTCCCAAGATGCA 1682  
QY 1682 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGGCGCAAAAAATCTTTTCTTG 1741  
Db 1683 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGGCGCAAAAAATCTTTTCTTG 1742

QY 1742 GGAGTCTACAAATTTGSAATGAAAAAACCCAGACATCAGATGTTTTTATTTATATAT 1801  
Db 1743 GGAGTCTACAAATTTGSAATGAAAAAACCCAGACATCAGATGTTTTTATTTATATAT 1802  
QY 1802 TATTTATGAAGGTGTGATACCATTTATCAATTTATGTGAAGGGACATGACACACCCAGCTT 1861  
Db 1803 TATTTATGAAGGTGTGATACCATTTATCAATTTATGTGAAGGGACATGACACACCCAGCTT 1862  
QY 1862 TGAAGGTGTGGGGGTGAAGACCTGAGCAGCCCCCACTGGGAACCAAGACTGACCTGGCCC 1921  
Db 1863 TGAAGGTGTGGGGGTGAAGACCTGAGCAGCCCCCACTGGGAACCAAGACTGACCTGGCCC 1922  
QY 1922 ATGGCTGTTTTCCCAAGATCAGTTCTGTGAAGGAAGGGCTGTGGCCCTGACTCCGCTGT 1981  
Db 1923 ATGGCTGTTTTCCCAAGATCAGTTCTGTGAAGGAAGGGCTGTGGCCCTGACTCCGCTGT 1982  
QY 1982 GTTCCGAGACACGTGTGACCGCAGCCCGCCCTGTGATGTTCTTGGCTGGGTCTGAG 2041  
Db 1983 GTTCCGAGACACGTGTGACCGCAGCCCGCCCTGTGATGTTCTTGGCTGGGTCTGAG 2042  
QY 2042 GTGTCTGTGAGACACCTGCTCCTCAACAGAGGTGAGCCACTTTCAGACTCCAGCT 2101  
Db 2043 GTGTCTGTGAGACACCTGCTCCTCAACAGAGGTGAGCCACTTTCAGACTCCAGCT 2102  
QY 2102 GAACATGGGAAACAACCTGAAGAGGAGGAGGCTCCCGGTCAAGGAGGCTGTGCTGTC 2161  
Db 2103 GAACATGGGAAACAACCTGAAGAGGAGGAGGCTCCCGGTCAAGGAGGCTGTGCTGTC 2162  
QY 2162 TGGCTTCCCATGACCACTCTCTCTGTGTAATATTACTGTTGAATCTGAGACAGATTG 2221  
Db 2163 TGGCTTCCCATGACCACTCTCTCTGTGTAATATTACTGTTGAATCTGAGACAGATTG 2222  
QY 2222 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCTTTT 2281  
Db 2223 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCTTTT 2282  
QY 2282 TTCCCACTCCCGAGAGCTGCTAAGCTATTCACAGGAGGCTTCCGCTTGGAGAGGTA 2341  
Db 2283 TTCCCACTCCCGAGAGCTGCTAAGCTATTCACAGGAGGCTTCCGCTTGGAGAGGTA 2342  
QY 2342 GTGTAACTCCGATCCAGAGGCTTGAAGCTTGAACAGAGTGTGCTTCCGTGTGG 2401  
Db 2343 GTGTAACTCCGATCCAGAGGCTTGAAGCTTGAACAGAGTGTGCTTCCGTGTGG 2402  
QY 2402 AAAAGCCCTGCAACAGGGCTTGAAGTGTGCTTGTCTTCCAGCCCTGCAACACC 2461  
Db 2403 AAAAGCCCTGCAACAGGGCTTGAAGTGTGCTTGTCTTCCAGCCCTGCAACACC 2462  
QY 2462 CACTTCAAGTGTGAGTGTGGGCAAGGCACTCAAGAGACAGGACAGGCGCTTGGC 2521  
Db 2463 CACTTCAAGTGTGAGTGTGGGCAAGGCACTCAAGAGACAGGACAGGCGCTTGGC 2522  
QY 2522 AAGACATCAGACACACCAACCAAGGCTGGAACCCAGGCGCGGCTGTATCCAG 2581  
Db 2523 AAGACATCAGACACACCAACCAAGGCTGGAACCCAGGCGCGGCTGTGTACAG 2582  
QY 2582 CAGGTGACCTGACAGTCCCGCTCTGTGAGGTCAAGGTCTCTCAAGAAACCAAGGCG 2641  
Db 2583 CAGGTGACCTGACAGTCCCGCTCTGTGAGGTCAAGGTCTCTCAAGAAACCAAGGCG 2642  
QY 2642 CTGTGCTCGAGGCTTCTCTGAGACCTTCTCTCAAGGAGCCACTTGGGATGCAAGATG 2701  
Db 2643 CTGTGCTCGAGGCTTCTCTGAGACCTTCTCTCAAGGAGCCACTTGGGATGCAAGATG 2702  
QY 2702 CAGCGAGCTAGAGACCCCTCCACGAGCTGAGACTTCCGCTGCAAGTAAAGTTACGTAGGC 2761  
Db 2703 CAGCGAGCTAGAGACCCCTCCACGAGCTGAGACTTCCGCTGCAAGTAAAGTTACGTAGGC 2762  
QY 2762 CTGTCTCTGGGGCTGGAAGTGTGAGCCATAGTTGCTTTGCTGACCCCTTCCGAGCA 2821  
Db 2763 CTGTCTCTGGGGCTGGAAGTGTGAGCCATAGTTGCTTTGCTGACCCCTTCCGAGCA 2822

QY	2822	GGGCGGACAGATGCTGCTGAGACAGCTGGGCGGGGGGGCCCAAGCTGGCGGGCTC	2881
Db	2823	GGCGCGCACAGTGGTGGCTAAGACAGCTGGCGGGGGGGCCCAAGCTGGCGGGCTC	2882
QY	2882	CAGCCCAACCCACAGCTGTTGCTGAAGTCAGGACCAACCTCCCGACAGCTGGTATCTGAATA	2941
Db	2883	CAGCCCAACCCACAGCTGTTGCTGAAGTCAGGACCTCCCGACAGACTGGTATCTGAATA	2942
QY	2942	ACGGCTAAGAACCTCTCTCTCTGGTTTGGAAAAGCACTTCGGGTTGTCCAAATTCGTAA	3001
Db	2943	ACGGCTAAGAACCTCTCTCTCTGGTTTGGAAAAGCACTTCGGGTTGTCCAAATTCGTAA	3002
QY	3002	CATTCATCTCATTTTTTAAAAAGGTTTCTCTGACGGCCCCAGCGCCCGCGGTG	3061
Db	3003	CATTCATCTCATTTTTT-AAAAAGGTTTCTCTGACGGCCCCAGCGCCCGCGGTG	3061
QY	3062	AGCGTCGTTGTCATGAGCTCGGGCGCCGGGCTTCCCGTGGGCTCTGCGCAGATGCTT	3121
Db	3062	AGCGTCGTTGTCATGAGCTCGGGCGCCGGGCTTCCCGTGGGCTCTGCGCAGATGCTT	3121
QY	3122	CTGGGCAACCCATCTCTCTGCGTTTCAATTTGACGTGACGTGTAACAAGGCACTCACAA	3181
Db	3122	CTGGGCAACCCATCTCTCTGCGTTTCAATTTGACGTGACGTGTAACAAGGCACTCACAA	3181
QY	3182	TAAACCTTTCCTGAAGC	3199
Db	3182	TAAACCTTTCCTGAAGC	3199

RESULT 4	AK001599	LOCUS	DEFINITION
	AK001599	3199 bp	mRNA
			linear
			Homo sapiens cDNA FLJ10737 f18, clone NT2RP3001426, weakly similar to DNAP PROTEIN.
			PRI 01-AUG-2002

ACCESSION AK001599  
VERSION AK001599.1 GI:7022951  
KEYWORDS oligo capping, fls (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

ORGANISM  
*Homo sapiens*,  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1. Isegai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Maesho, Y., Ninomiya, K. and Iwayanagi, T.

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 3199)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute.

**COMMENT**

Genomics Laboratory, 1532-3 Yana, Katsurazaki, Chiba 229-0812, Japan (E-mail: genomics@ntt.co.jp, Tel: 81-438-52-1975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES	SOURCE
location/Qualifiers	1. 3199
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="NT2RP3001426"	
/cell_line="NT2"	
/cell_type="teratocarcinoma"	
/clone_1ib="NT2RP3"	

```

CDS
/note="Cloning vector: pMT18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
16..1695
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91780.1"
/db_xref="GI:7022952"
/translation="MATATLSSEELDNEEDYSLINVRREASSEELKAAVRLCMHYHDD
KXHDPELKSOERLFFNLVHOAYEVLSPQFRAIYDIYGRKLEMEHGEVVERKRTTPAE
IRREPERLQREEREERLRQQRTPNPKGITSVGDATDLDPRADDEEYDVSGSPFOEIN
KMIISQIEAPLTATPDAILSGSLSTONGSGSINLRAVLVRSAKMGKLEFPGAGL
QGLPFGKLTFRNLTPRCFTVTNCAOLGSSSGIIRGLTVALRNLIDKTMWYLIQWRGI
OSAMNTEIVRDTKTSHTFTALOGIHPSPALIYIOHKFODDODRVRVSGAKGPGTV
VEYGAEKRTISRHSHVLAANVGVPCGSLINLRASQTPPEPHITDOLIPAMFPA
TVGHLVYVYFAMERKLIITKPYLRQAKERELEQRSASATDVTQKQBAASAVRLMSQRA
RIILEAESRRMGLIIVNAWGRFVNDKSRKSEKKAVIDVTLPLQCLVDSKSLIILEASK
AGLPGFYDPCVGEENKLVLYLPGRGVLAQWVLDSEALRIPIKQSHRIDTGG"
ORIGIN

```

Query Match	98.7%	Score 3171.6	DB 9	Length 3199
Best Local Similarity	99.7%	Pred. No. 0		
Matches 3188	Conservative	0	Mismatches 9	Indels 1
			Gaps	1
QY	2	AAAGGTGCAGAAAGATGCGACGAGCTTTGACGCGAGAGAGACTGGACAATGAAAGCTATTTA	61	
Db	3	AAAGGTGCAGAAAGATGCGACGAGCTTTGACGCGAGAGAGACTGGACAATGAAAGCTATTTA	62	
QY	62	CTGGTTCTGAAAGCTGCGCAGGAGAGGCTTTCTGAAGAGCTGAAGCTGCTACCGGAG	121	
Db	63	CTCGTTCTGAAAGCTGCGCAGGAGAGGCTTTCTGAAGAGCTGAAGCTGCTACCGGAG	122	
QY	122	GCTGTGATGCTCTACCATCCAGCACAAGACCCAGAGCTCAAGTCAACAGCGGGA	181	
Db	123	GCTGTGATGCTCTACCATCCAGCACAAGAGCTCAAGTCAACAGCGGGA	182	
QY	182	ACGACTGTTTAACTTGTTCACCGCTTATGAAGTCTTAGTGAACCCCAACAGAGGC	241	
Db	183	ACGACTGTTTAACTTGTTCACCGAGCTTACGAAGTCTTAGTGAACCCCAACAGAGGC	242	
QY	242	CATCTATGATATATATGGAAGAGAGACTGGAAATGAAAGATGGAGGTTGTGAAAG	301	
Db	243	CATCTATGATATATATGGAAGAGAGACTGGAAATGAAAGATGGAGGTTGTGAAAG	302	
QY	302	GAGAGAAACCCCTGCTGAATTTGAGAGAGCTTTGACCGCTGACAGAGAGAGAGAGA	361	
Db	303	GAGAGAAACCCCTGCTGAATTTGAGAGAGCTTTGACCGCTGACAGAGAGAGAGAGA	362	
QY	362	GAGAGATTGACAGAGGGAACCAATCCCAAGGGAACATCAGCTTGGAGTAGATGCCAC	421	
Db	363	GAGAGATTGACAGAGGGAACCAATCCCAAGGGAACATCAGCTTGGAGTAGATGCCAC	422	
QY	422	CGACCTTTTGTGCTATGATGAGAGATGAAGATGTGTCCGCAAGTACCTTCCGCA	481	
Db	423	CGACCTTTTGTGCTATGATGAGAGATGAAGATGTGTCCGCAAGTACCTTCCGCA	482	
QY	482	GATTGAAATTAATAAATGCAATATCCCAAGTCCATTGAGAGCAACCTTTGACAGGACAGA	541	
Db	483	GATTGAAATTAATAAATGCAATATCCCAAGTCCATTGAGAGCAACCTTTGACAGGACAGA	542	
QY	542	CACAGCATTCTCTGGAAGCTCTTCAACCCAGAAATGGAATGGAAGAGGTTCCATTAA	601	
Db	543	CACAGCATTCTCTGGAAGCTCTTCAACCCAGAAATGGAATGGAAGAGGTTCCATTAA	602	
QY	602	CTTTGCGCTCAGACGAGTAACTTTGCGCAAAAGGATGGGAGAGTTGGAAATTTGGAGCTGG	661	
Db	603	CTTTGCGCTCAGACGAGTAACTTTGCGCAAAAGGATGGGAGAGATTTGGAGCTGG	662	
QY	662	AGACTTACAGAGGGCTTTGTTCGCTCTCAAGCTGTTCCGTAAATCTCAACCAAGATCTT	721	
Db	663	AGACTTACAGAGGGCTTTGTTCGCTCTCAAGCTGTTCCGTAAATCTCAACCAAGATCTT	722	
QY	722	TGTGACAAACAACGTGCTCTGACAGTTTATCCCGTGAAATCGAACCCGAGCTGACAC	781	

|||||  
Db 723 TGTGACAACTGCTCTGACAGTTTTCATCCCGGAAATCCGACCCCGGCTGACCA 782  
QY 782 TGTCTAGAGCTGGAACTAGACAAAGAACACGCGGGCTACCTGACAGCGGATAGGGGAT 841  
Db 783 TGTCTAGAGCTGGAACTAGACAAAGAACACGCGGGCTACCTGACAGCGGATAGGGGAT 842  
QY 842 CCAAGTCAAGCCATGAACCTAGACATCGTCCGAGACACTAAACCAAGCCACTTCACTGTGC 901  
Db 843 CCAAGTCAAGCCATGAACCTAGACATCGTCCGAGACACTAAACCAAGCCACTTCACTGTGC 902  
QY 902 CTTGCAAGCTGGGAATCCCTCACTCTCTTGGCACTGATCAGTATACAGCAAAATTTCCAAG 961  
Db 903 CTTGCAAGCTGGGAATCCCTCACTCTCTTGGCACTGATCAGTATACAGCAAAATTTCCAAG 962  
QY 962 TGAAGATCAGACTGCTGTAAGAGATCCCTCAAGAGAGGCTCTTGGGAGCGGTGGTGA 1021  
Db 963 TGAAGATCAGACTGCTGTAAGAGATCCCTCAAGAGAGGCTCTTGGGAGCGGTGGTGA 1022  
QY 1022 GTACGAGCTGAGAGAGAGATCTCCAGGCAAGCGTTTGGGTGACAGCTGTCAAGCGTTG 1081  
Db 1023 GTACGAGCTGAGAGAGAGATCTCCAGGCAAGCGTTTGGGTGACAGCTGTCAAGCGTTG 1082  
QY 1082 AGTTTCCACAGGGGCTTTCTCTCAAAGTCAAGCTCAAGGGCCAGTCAACATATCTTT 1141  
Db 1083 AGTTTCCACAGGGGCTTTCTCTCAAAGTCAAGCTCAAGGGCCAGTCAACATATCTTT 1142  
QY 1142 CCGTATTCATTTGACGAGACAGGCTTCCGACGAGGCAATGTTATGCGACCGTGGAGGC 1201  
Db 1143 CCGTATTCATTTGACGAGACAGGCTTCCGACGAGGCAATGTTATGCGACCGTGGAGGC 1202  
QY 1202 TCTAGTGTCTACTTTGCTCCATGACCGTCTGATCATCAACATCACTCAAGGCTCAGAA 1261  
Db 1203 TCTAGTGTCTACTTTGCTCCATGACCGTCTGATCATCAACATCACTCAAGGCTCAGAA 1262  
QY 1262 AGAAGAGGAATTTGAG 1321  
Db 1263 AGAAGAGGAATTTGAG 1322  
QY 1322 AGAGGCGGAGTCCGCTGTCCGCTGATGACGAGATCTGTCCGAGAGATTTAGAGGAG 1381  
Db 1323 AGAGGCGGAGTCCGCTGTCCGCTGATGACGAGATCTGTCCGAGAGATTTAGAGGAG 1382  
QY 1382 AGAGTCCAGATGAGGCTCATCATCTGTCATGCTGCTGTAAGAGAGTTTGTCAATGACA 1441  
Db 1383 AGAGTCCAGATGAGGCTCATCATCTGTCATGCTGCTGTAAGAGAGTTTGTCAATGACA 1442  
QY 1442 GAGCAG 1501  
Db 1443 GAGCAG 1502  
QY 1502 GAGAGACTCGAGCTCATCTCAAGAGGCTCCAGAGGCTGGGCTGCTGGCTTTATGA 1561  
Db 1503 GAGAGACTCGAGCTCATCTCAAGAGGCTCCAGAGGCTGGGCTGCTGGCTTTATGA 1562  
QY 1562 CCGGTGTGTGGGAG 1621  
Db 1563 CCGGTGTGTGGGAG 1622  
QY 1622 TCAAGT 1681  
Db 1623 TCAAGT 1682  
QY 1682 TACAGATGATTAAGTCCAG 1741  
Db 1683 TACAGATGATTAAGTCCAG 1742  
QY 1742 GAGAGTCAAAATTTGAG 1801  
Db 1743 GAGAGTCAAAATTTGAG 1802  
QY 1802 TATTTATGAG 1861

Db 1803 TATTTATGAG 1862  
QY 1862 TGAAGGTGTGGAGGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1921  
Db 1863 TGAAGGTGTGGAGGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1922  
QY 1922 ATGGCTGTTTTCCAAAGATCAAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1981  
Db 1923 ATGGCTGTTTTCCAAAGATCAAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1982  
QY 1982 GTTCCGAGACACGTGTCAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2041  
Db 1983 GTTCCGAGACACGTGTCAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2042  
QY 2042 GTGTCTGTGAGACACCTTCCCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2101  
Db 2043 GTGTCTGTGAGACACCTTCCCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2102  
QY 2102 GAAATGGGAAACAACTGTAAG 2161  
Db 2103 GAAATGGGAAACAACTGTAAG 2162  
QY 2162 TGGCTTCCATGACCACT 2221  
Db 2163 TGGCTTCCATGACCACT 2222  
QY 2222 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2281  
Db 2223 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2282  
QY 2282 TTCCCACTCCCAAGAGCTGCAAGTCAATTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2341  
Db 2283 TTCCCACTCCCAAGAGCTGCTCAAAATCATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2342  
QY 2342 GTGTAACTCCCAAG 2401  
Db 2343 GTGTAACTCCCAAG 2402  
QY 2402 AAAAGGCTTCCCAAG 2461  
Db 2403 AAAAGGCTTCCCAAG 2462  
QY 2462 CACTTCAAGAGTGAAGTGTGGGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2521  
Db 2463 CACTTCAAGAGTGAAGTGTGGGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2522  
QY 2522 AAGACATCAAGACACCAACCAAG 2581  
Db 2523 AAGACATCAAGACACCAACCAAG 2582  
QY 2582 CAGGTGACACTGAGTCCCGCTCCGCTGAGAGTCCAGAGTCTCTCAAGAGAACAGAGAGAG 2641  
Db 2583 CAGGTGACACTGAGTCCCGCTCCGCTGAGAGTCCAGAGTCTCTCAAGAGAACAGAGAGAG 2642  
QY 2642 CTGTGCTCCGAGAGCTTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2701  
Db 2643 CTGTGCTCCGAGAGCTTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2702  
QY 2702 CAGCGAGTGAAG 2761  
Db 2703 CAGCGAGTGAAG 2762  
QY 2762 CTGTCTCTGAG 2821  
Db 2763 CTGTCTCTGAG 2822  
QY 2822 GCGCGGACAGAGT 2881  
Db 2823 GCGCGGACAGAGT 2882  
QY 2882 CAGCCCAACCAAGAGT 2941  
Db 2883 CAGCCCAACCAAGAGT 2942

OY	2942	ACGGTAGAAGACCTCCCTCCTCGGTTTGTAAAGACAGCTTCGGGTGTGCCAATTCGTGA	3001
Db	2943	ACGGTAAGAACCCTCCTCCTCGGTTTGTAAAGACAATTGGGTGCCAATTCGTGA	3002
OY	3002	CATTCATCTCCATTTTTTAAAAAGGTTTCTCTGACGGCCCCACGGCCCCAGCGCGGT	3061
Db	3003	CATTCATCTCCATTTTTT-AAAAGGTTTCTCTGACGGCCCCACGGCCCCAGCGCGGT	3061
OY	3062	AGCGTCGTGTGCATGAGCCTTGCGGCCCGGGCCTTCCCCTGCGCCCTTCGCGCAGTGTCT	3121
Db	3062	AGCGTCGTGTGCATGAGCCTTGCGGCCCGGGCCTTCCCCTGCGCCCTTCGCGCAGTGTCT	3121
OY	3122	CTGGCACCCACCTCCCTGCGTTTCAATTGGACGTGACCTGTACAGAGGACTACACCAA	3181
Db	3122	CTGGCACCCACCTCCCTGCGTTTCAATTGGACGTGACCTGTACAGAGGACTACACCAA	3181
OY	3182	TAAACCTTCTCGAAGC	3199
Db	3182	TAAACCTTCTCGAAGC	3199
RESULT 5	HSMB05121	3178 bp	mRNA linear PRI 12-JUN-2002
LOCUS	HSMB05121		
DEFINITION	Homo sapiens mRNA; cDNA DKFZp727C181 (from clone DKFZp727C181).		
ACCESSION	AL833841		
VERSION	AL833841.1	GI:21739308	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 3178)		
JOURNAL	Poustka,A., Klein,W., Mewes,H.W., Gassenhuber,J. and Wiemann,S. Direct Submission Submitted (10-JUL-2002) GSF, Institut fuer Bioinformatik MIPS, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp727C181) is available at the RZPD in Berlin. Please contact the RZPD: Resourcenentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/. Location/Qualifiers		
COMMENT	1..3178		
FEATURES	/organism="Homo sapiens"		
source	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/map="tp36.21-36.33"		
	/clone="DKFZp727C181"		
	/tissue_type="breast cancer"		
	/clone_id="727 (synonym: hmcf1). Vector pSport1; host DH10B; sites NotI + SalI"		
	/dev_stage="adult"		
	1..3178		
	/gene="DKFZp727C181"		
	<1..1660		
	/gene="DKFZp727C181"		
	/note="weak similarity to putative DNAJ protein (A. thaliana)"		
	/codon_start=2		
	/product="hypothetical protein"		
	/protein_id="CAD38701.1"		
	/db_xref="gi:21739309"		
	/db_xref="SPRTEMBL:OBNDM2"		
	/translation="EELDNEDYISLAWVRKASEELKAAYRIQLMYPDKRIDPEL KSQERFNVLVHQAEVALSDQTSAIDYIGKRGLKEGWEEVERRTPAIEEFER LQRERRERLDQORTNPNGKITISVGDAIDLPGRKLEDEWGSSEFPDIEINKHISQS		
gene			
CDS			

Query Match	98.3%	Score 3157.4	DB 9	Length 3178
Best Local Similarity	99.8%	Pred. No. 0		
Matches 3172	Conservative 0	Mismatches 6	Indels 1	Gaps 1
QY	35	GAGAGAGCTGACCAATGAAAGCTATATACCTGTTCTGAACGTGGCAGAGAGGCTCTTC	94	
Db	1	GGAGAGCTGACCAATGAAAGCTATATACCTGTTCTGAACGTGGCAGAGAGGCTCTTC	60	
QY	95	TGAAGAGCTGAAAGCTGCTACCGAGAGCTCTGTATGCTTACCATCAGACAGACAG	154	
Db	61	TGAAGAGCTGAAAGCTGCTACCGAGAGCTCTGTATGCTTACCATCAGACAGACAG	120	
QY	155	AGACCCAGAGCTCAAGTCAACAGGCGGAAACGACTGTTAACCTTGTCCACAGGCTTATGA	214	
Db	121	AGACCCAGAGCTCAAGTCAACAGGCGGAAACGACTGTTAACCTTGTCCACAGGCTTATGA	180	
QY	215	AGTGTATTGTAACCCCAACCCAGGCGCATTTATGATATATATGGAAGAGAGACTGGA	274	
Db	181	AGTGTATTGTAACCCCAACCCAGGCGCATTTATGATATATATGGAAGAGAGACTGGA	240	
QY	275	AATGGAAGATGGAGGTTGTGTGAAGAAGAGAGAACCCCTGCTGAATTGAGAGAGTT	334	
Db	241	AATGGAAGATGGAGGTTGTGTGAAGAAGAGAGAACCCCTGCTGAATTGAGAGAGTT	300	
QY	335	TGAAGCGCTGCAGAGAGAGAGAGAGAGAGATTGACAGAGCAACCATCCCAAGG	394	
Db	301	TGAAGCGCTGCAGAGAGAGAGAGAGAGAGATTGACAGAGCAACCATCCCAAGG	360	
QY	395	AACGATCAGCGTTGAGATGATGCCACCGACCTTTTGATGCTATGATGAGAGATGATGA	454	
Db	361	AACGATCAGCGTTGAGATGATGCCACCGACCTTTTGATGCTATGATGAGAGATGATGA	420	
QY	455	AGATGTGCGCGGACGTAGCTTTCCGCGAGTTGAATAATTAATGACATATCCCACTC	514	
Db	421	AGATGTGCGCGGACGTAGCTTTCCGCGAGTTGAATAATTAATTAATGACATATCCCACTC	480	
QY	515	CATTGAGGCACTCTTGACAGGCAAGACACAGACCATCTCTGGAAGCTCTCAACCCA	574	
Db	481	CATTGAGGCACTCTTGACAGGCAAGACACAGACCATCTCTCTGGAAGCTCTCAACCCA	540	
QY	575	GAATGGAATGAGAGAGGTTTCATTAACTTTGCGCTCAGACGAGTAACTTGGCAAAAGG	634	
Db	541	GAATGGAATGAGAGAGGTTTCATTAACTTTGCGCTCAGACGAGTAACTTGGCAAAAGG	600	
QY	635	ATGGGAGAGTTTGAATTTTGAAGCTGGAAGACTTACAGGGGCTTTTGTTCGATCTCAAGCT	694	
Db	601	ATGGGAGAGTTTGAATTTTGAAGCTGGAAGACTTACAGGGGCTTTTGTTCGATCTCAAGCT	660	
QY	695	GTTCCGTTATCTCACCAAGATGCTTTGTGACAAACAACTGTGCTCGAGTTTTCATC	754	
Db	661	GTTCCGTTATCTCACCAAGATGCTTTGTGACAAACAACTGTGCTCGAGTTTTCATC	720	
QY	755	CCGTGGAATCCGACCCGGCTGACCACTGTCTTACGCTCGGAACCTTAGCAAGACCGT	814	
Db	721	CCGTGGAATCCGACCCGGCTGACCACTGTCTTACGCTCGGAACCTTAGCAAGACCGT	780	
QY	815	GGGCTACTTGAGTGGGATGGGGTATTCAGTCAAGCCATGAACACTAGCATCGTCCGAGA	874	
Db	781	GGGCTACTTGAGTGGGATGGGGTATTCAGTCAAGCCATGAACACTAGCATCGTCCGAGA	840	

QY 875 CACTAAAAACGACCTTCACTGTGGCCCTGACGCTGGGAATCCCTCACTCTTTGCACT 934  
Db 841 CACTAAAAACGACCACTTCACTGTGGCCCTGACGCTGGGAATCCCTCACTCTTTGCACT 900  
QY 935 GATCAGCTATCAGACAAATTTCCAAAGATGACGATCAGATCTCGTGTGAAGAAGATCCCTCA 994  
Db 901 GATCAGCTATCAGACAAATTTCCAAAGATGACGATCAGATCTCGTGTGAAGAAGATCCCTCA 960  
QY 995 AGCAGCTTCTTTGGAGCGGTGTGTGAGTACGAGAGCTGAGAGAAAGATCTCCAGGACAG 1054  
Db 961 AGCAGCTTCTTTGGAGCGGTGTGTGAGTACGAGAGCTGAGAGAAAGATCTCCAGGACAG 1020  
QY 1055 CGTTTGGGTGACAGCTGTACGCTTGGAGTTTCCACAGGCGCTTCTCTCAAGTCAAGCT 1114  
Db 1021 CGTTTGGGTGACAGCTGTACGCTTGGAGTTTCCACAGGCGCTTCTCTCAAGTCAAGCT 1080  
QY 1115 CAACAGGCGCAGTCAAGATCTTCTTCCCTATTCATTTGACGGAACGAGTCTTGCCAG 1174  
Db 1081 CAACAGGCGCAGTCAAGATCTTCTTCCCTATTCATTTGACGGAACGAGTCTTGCCAG 1140  
QY 1175 CGCATGTTCTATGCCACCGTGGGGCTCTAGTGTCTACTTTGGCATGACCGTCTGAT 1234  
Db 1141 CGCATGTTCTATGCCACCGTGGGGCTCTAGTGTCTACTTTGGCATGACCGTCTGAT 1200  
QY 1235 CATCAAAACATACCTCAGGCTCAGAAAAGAAAGAAATTTGAGAGACAGAGGAAAGCGC 1294  
Db 1201 CATCAAAACATACCTCAGGCTCAGAAAAGAAAGAAATTTGAGAGACAGAGGAAAGCGC 1260  
QY 1295 CGCCACCGATGTGTGAGAGAAAGAGAGCGAGATCCGCTGTCCGGTGTATGACAGA 1354  
Db 1261 CGCCACCGATGTGTGAGAGAAAGAGAGCGAGATCCGCTGTCCGGTGTATGACAGA 1320  
QY 1355 ATCTGTCCGAAGATTAATTAGGAGAGAGATCCAGATGGGCTCATCTCATATGC 1414  
Db 1321 ATCTGTCCGAAGATTAATTAGGAGAGAGATCCAGATGGGCTCATCTCATATGC 1380  
QY 1415 CTGGATCGGAGATTTGTCAATGACAAAGACAGAGAGCGAGAGAGTGAAGTGAATTA 1474  
Db 1381 CTGGATCGGAGATTTGTCAATGACAAAGACAGAGAGCGAGAGAGTGAAGTGAATTA 1440  
QY 1475 CGTGAAGTGCCTCGTAGGCGCTGTGAGAGACTCGAAGCTCATCTTCAGAGGCTTC 1534  
Db 1441 CGTGAAGTGCCTCGTAGGCGCTGTGAGAGACTCGAAGCTCATCTTCAGAGGCTTC 1500  
QY 1535 CAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGAGAGAGAGAACTGGAAGT 1594  
Db 1501 CAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGAGAGAGAACTGGAAGT 1560  
QY 1595 GCTCTATCAGTTCCGGGGCTCTCGATCAGGTGATGTGCTGACAGTGAAGCCCTCCG 1654  
Db 1561 GCTCTATCAGTTCCGGGGCTCTCGATCAGGTGATGTGCTGACAGTGAAGCCCTCCG 1620  
QY 1655 GATACCAAGCAGTCCCAAGATCGATACAGATGATAAATGCGCAAGAACCAAGATTTT 1714  
Db 1621 GATACCAAGCAGTCCCAAGATCGATACAGATGATAAATGCGCAAGAACCAAGATTTT 1680  
QY 1715 TAAAAAGCCCAAAAAATCTTTTCTGGAGTCTACAATTTTGGAAATGAAAAACCCAG 1774  
Db 1681 TAAAAAGCCCAAAAAATCTTTTCTGGAGTCTACAATTTTGGAAATGAAAAACCCAG 1740  
QY 1775 ACATCAGATGTTTTATTTATTAATTAATTAATTAAGAGTGTGACATTAATTAATGT 1834  
Db 1741 ACATCAGATGTTTTATTTATATTAATTAATTAAGAGTGTGACATTAATTAATGT 1800  
QY 1835 GAAGGAGCATGAGACCCCAAGCTTTTGAAGGTGTGAGGAGTGAAGATGAGGACAGCC 1894  
Db 1801 GAAGGAGCATGAGACCCCAAGCTTTTGAAGGTGTGAGGAGTGAAGATGAGGACAGCC 1860  
QY 1895 ACTGGGAACGAGCTGAGCTGGCCATGCTGTTTCCCAAGATCAGTCTGAGG 1954  
Db 1861 ACTGGGAACGAGCTGAGCTGGCCATGCTGTTTCCCAAGATCAGTCTGAGG 1920  
QY 1955 GAAGGCTCTGGCCCTGATCTCCGTGTGTCGAGACACGTCGTGACCGCAGCCCGCG 2014

Db 1921 GAAGGCTCTGGCCCTGATCTCCGTGTGTCGAGACACGTCGTGACCGCAGCCCGCG 1980  
QY 2015 CCTGTAGTTCTTGGCTGGGCTGGAGGTGTCTGTGAGACACCTGGCCCTCACAAGGA 2074  
Db 1981 CCTGTAGTTCTTGGCTGGGCTGGAGGTGTCTGTGAGACACCTGGCCCTCACAAGGA 2040  
QY 2075 GCGTGAACCATTTCTGACATGCCGTGAACATGGGAAACAACTGAAAAAGCAGGCGC 2134  
Db 2041 GCGTGAACCATTTCTGACATGCCGTGAACATGGGAAACAACTGAAAAAGCAGGCGC 2100  
QY 2135 CTCGCCGTGAGGAGCTCTGTGTGTGCTGCTTCCATGACCACTCTCTCTGTAAT 2194  
Db 2101 CTCGCCGTGAGGAGCTCTGTGTGTGCTGCTTCCATGACCACTCTCTCTGTAAT 2160  
QY 2195 ATTTAGCTTGAATCTGAGACGATTTGGCGGTTTAAACCTGTTTATCTGAGAAACA 2254  
Db 2161 ATTTAGCTTGAATCTGAGACGATTTGGCGGTTTAAACCTGTTTATCTGAGAAACA 2220  
QY 2255 AACGGTTTGAATAATGATGATCTTTTCCCACTCCCAAGCTGCTCAAGTCAATCCA 2314  
Db 2221 AACGGTTTGAATAATGATGATCTTTTCCCACTCCCAAGCTGCTCAATCAATCCA 2280  
QY 2315 CGGCCCCCTCGGCTTGGGACAGGGTATGTAACTCCGATCCAGGGCTTGGCCCTGAC 2374  
Db 2281 CGGCCCCCTCGGCTTGGGACAGGGTATGTAACTCCGATCCAGGGCTTGGCCCTGAC 2340  
QY 2375 ACAGGTGCTTCCCGATCCCGGTGGGAAAGCCCTGACACAGCGGGCTGAGCTGAC 2434  
Db 2341 ACAGGTGCTTCCCGATCCCGGTGGGAAAGCCCTGACACAGCGGGCTTGAAGTGA 2400  
QY 2435 CTGTGTCTCTCCACCGCTGACCAACCACTTCAGAGTGCAGTGTGGGCAAGGCAAC 2494  
Db 2401 CTGTGTCTCTCCACCGCTGACCAACCACTTCAGAGTGCAGTGTGGGCAAGGCAAC 2460  
QY 2495 TCAAGAGACAGACCAAGGCGCTTGGCAAGATCAGACACCAACCAAGGCGTGG 2554  
Db 2461 TCAAGAGACAGACCAAGGCGCTTGGCAAGATCAGACACCAACCAAGGCGTGG 2520  
QY 2555 ACCCCAGGCGCGGCGCTGGTAAACCAAGAGTGGCACTGACCTCCGCTCTGCAAGT 2614  
Db 2521 ACCCCAGGCGCGGCGCTGGTAAACCAAGAGTGGCACTGACCTCCGCTCTGCAAGT 2580  
QY 2615 CCAGCTCTCTCAGAGAACCAAGGCGTGTCTCCGAGCCCTTCTTCAACCTTCTCT 2674  
Db 2581 CCAGCTCTCTCAGAGAACCAAGGCGTGTCTCCGAGCCCTTCTTCAACCTTCTCT 2640  
QY 2675 CCACTGCCCCATTGGGATGCAAGATGACGCGGAGTGAAGTCCCTCCAAGGCTGAC 2734  
Db 2641 CCACTGCCCCATTGGGATGCAAGATGACGCGGAGTGAAGTCCCTCCAAGGCTGAC 2700  
QY 2735 CTGCGGTGAGTAAAGTTACGTGAGGCGTGTCTCGGGGCGCTGGAAGTGGAGCATCA 2794  
Db 2701 CTGCGGTGAGTAAAGTTACGTGAGGCGTGTCTCGGGGCGCTGGAAGTGGAGCATCA 2760  
QY 2795 GTTGTCTTGTGACCCCTCGAGACAGCGCCGACAGGTGTGTGCTGAGACGTGAGC 2854  
Db 2761 GTTGTCTTGTGACCCCTCGAGACAGCGCCGACAGGTGTGTGCTGAGACGTGAGC 2820  
QY 2855 CGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCAAGCTGTTGCTGAAGTCAAGCC 2914  
Db 2821 CGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCAAGCTGTTGCTGAAGTCAAGCC 2880  
QY 2915 AACCTCCCAAGACGTGTATCTGAGTACGCTGAAGAACTGCTCTCTGTTTGA 2974  
Db 2881 TCCCTCCCAAGACGTGTATCTGAGTACGCTGAAGAACTGCTCTCTGTTTGA 2940  
QY 2975 AGCAGTTCCGGTGTGTCATTTCTGTAACTTCATCTCAATTTTAAAAAGTTTCTCT 3034  
Db 2941 AGCAGTTCCGGTGTGTCATTTCTGTAACTTCATCTCAATTTTAAAAAGTTTCTCT 2999  
QY 3035 GACGCGCCCAAGGCGCGGAGCGCGGTGAGCTGTGTGCTGAGCTGTGGCCCGGCGCT 3094

Db 3000 GACGGCCCCAGCGCCGAGCGCGGTGAGCGTGTGTTGATGAGCTTGAGCCCCGGGCT 3059  
Qy 3095 TCCCGTGGCCTTCTGCCAGGTGTTCTGGGCAACCATCTCTCCGTTTCATTGTCAGT 3154  
Db 3060 TCCCGTGGCCTTCTGCCAGGTGTTCTGGGCAACCATCTCTCCGTTTCATTGTCAGT 3119  
Qy 3155 CGACTGTACAGAGGCACTCACACATTAACCTTCTGGAAGCAGAAAAA 3213  
Db 3120 CGACTGTACAGAGGCACTCACACATTAACCTTCTGGAAGCAGAAAAA 3178

RESULT 6  
AK025491 2064 bp mRNA linear PRI 13-SEP-2003  
LOCUS Homo sapiens cDNA: FLJ21838 f1s, clone HEP01752, highly similar to  
DEFINITION AK025491  
ACCESSION AK025491  
VERSION AK025491.1 GI:10438023  
KEYWORDS oligo capping; f1s (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1  
Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,  
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Ohyashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2064)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Ohayashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.

TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: flicdn@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction: 5' - & 3' - end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source  
1..2064  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP01752"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_id="HEP"  
/note="Cloning vector pME18FLJ3"  
misc\_feature  
1..2064  
/note="highly similar to AF052158 Homo sapiens clone 24616  
mRNA sequence"

ORIGIN  
Query Match 60.6%; Score 1947; DB 9; Length 2064;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1972; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
Qy 1236 ATCAACCATACCTCAGGGCTCAGAAAGAAAGATTGAGAAAGAGGAAAGCCGC 1295  
Db 2 AACAAACATACCTCAGGGCTCAGAAAGAAAGATTGAGAAAGAGGAAAGCCGC 61  
Qy 1296 GCCACCATATGCTGCGAAGAAAGAGGCGAGTCCGCTGTCGGCTATGACGAA 1355  
Db 62 GCCACCGATGCTGCGAAGAAAGAGGCGAGTCCGCTGTCGGCTATGACGAA 121  
Qy 1356 TCTGTCCGAAGATTAATTGAGGCAAGAGTCCAGAAATGGGCTCATGTCATGCC 1415

Db 122 TCTGTCCGAAGATTAATTGAGGCAAGAGTCCAGAAATGGGCTCATATGTCATGCC 181  
Qy 1416 TGTACGCGAAAGTTTGTCAATGACAAGAGCAGAAAGCGAAGGTGAAGTATTGAC 1475  
Db 182 TGTACGCGAAAGTTTGTCAATGACAAGAGCAGAAAGCGAAGGTGAAGTATTGAC 241  
Qy 1476 GTGACTGTGCCCCCTGCAAGTCTGTGTGAAGGACTGCAAGTCTTCAAGAGGCTCC 1535  
Db 242 GTGACTGTGCCCCCTGCAAGTCTGTGTGAAGGACTGCAAGTCTTCAAGAGGCTCC 301  
Qy 1536 AAGCTGAGCTGCTGAGCTTTTATGACCCGCTGTGGGCGAAGAAAGCTGAAAGT 1595  
Db 302 AAGCTGAGCTGCTGAGCTTTTATGACCCGCTGTGTGGGCGAAGAAAGCTGAAAGT 361  
Qy 1596 CTCTATGATGTCGGGGCGTCTGTCATCAGGTGATGCTGTCAGTGTAGGCGCTCCG 1655  
Db 362 CTCTATGATGTCGGGGCGTCTGTCATCAGGTGATGCTGTCAGTGTAGGCGCTCCG 421  
Qy 1656 ATACCAAGCAGTCCACAGATGATACAGATGATTAACCTGCCAAGAACAGATTTT 1715  
Db 422 ATACCAAGCAGTCCACAGATGATACAGATGATTAACCTGCCAAGAACAGATTTT 481  
Qy 1716 AAAAGCCGCAAAAATCTTTTCTGGAGTCTACAAATTTGAAATGAAATCCAGA 1775  
Db 482 AAAAGCCGCAAAAATCTTTTCTGGAGTCTACAAATTTGAAATGAAATCCAGA 541  
Qy 1776 CATCAGATG-TTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1834  
Db 542 CATCAGATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 601  
Qy 1835 GAAGGACATGACACACCCAGTCTTGAAGGCTGCGGGGTGAGCTGAGGAGCGCC 1894  
Db 602 GAAGGACATGACACACCCAGTCTTGAAGGCTGCGGGGTGAGCTGAGGAGCGCC 661  
Qy 1895 ACTGGAACAGACATGACGCTGAGCCGATGAGCTGTTTCCAAAGATCAGTCTGAGG 1954  
Db 662 ACTGGAACAGACATGACGCTGAGCCGATGAGCTGTTTCCAAAGATCAGTCTGAGG 721  
Qy 1955 GAAGGCTCTGAGCCGCTGACCTGCTGATGCTCCGAGCAGAGTGTGACCGAGCCCG 2014  
Db 722 GAAGGCTCTGAGCCGCTGACCTGCTGATGCTCCGAGCAGAGTGTGACCGAGCCCG 781  
Qy 2015 CCTGTATGTTCTGAGCTGAGGCTGAGAGGTGTCTGTGAGACACCTGCTCAGCAG 2074  
Db 782 CCTGTATGTTCTGAGCTGAGGCTGAGAGGTGTCTGTGAGACACCTGCTCAGCAG 841  
Qy 2075 GCGTGAACCACTCTGAGTCCACGCTGAACATGAGGAAACAACCTGAAAGAGGCGAGG 2134  
Db 842 GCGTGAACCACTCTGAGTCCACGCTGAACATGAGGAAACAACCTGAAAGAGGCGAGG 901  
Qy 2135 CTCCCGTCAAGGAGCCTCTGCTGTGCTGCTCCATGACACCTCTCTCTGTAAT 2194  
Db 902 CTCCCGTCAAGGAGCCTCTGCTGTGCTGCTCCATGACACCTCTCTCTGTAAT 961  
Qy 2195 ATTATGCTGTAATCTGAGACAGATTGCGGGTTTATTAACCTGCTTTTATCTGAACA 2254  
Db 962 ATTATGCTGTAATCTGAGACAGATTGCGGGTTTATTAACCTGCTTTTATCTGAACA 1021  
Qy 2255 AAGCGTTTGAAGAAATTATGCTGCTTTTCCCACTCCGAGAGTGTCAAGTATTC 2314  
Db 1022 AAGCGTTTGAAGAAATTATGCTGCTTTTCCCACTCCGAGAGTGTCAAGTATTC 1081  
Qy 2315 CCGGCCCCCTCGCTTGGGACAGGGTATGTAATCCCGATCCAGAGGCTTAGCCCTGAC 2374  
Db 1082 CCGGCCCCCTCGCTTGGGACAGGGTATGTAATCCCGATCCAGAGGCTTAGCCCTGAC 1141  
Qy 2375 ACAGTGGCTTCCCGTATCCCGGTGGAAACGCTCTGCAACAGCGGGCTTGAAGTGGC 2434  
Db 1142 ACAGTGGCTTCCCGTATCCCGGTGGAAACGCTCTGCAACAGCGGGCTTGAAGTGGC 1201  
Qy 2435 CTGTGCTCTGCAACCGGCTGACCAACCTCAGAGTGCAGTGTGGGCAAGGCGAGC 2494



```

Db      1202 CTGTGTCCCCCAACCGCTGACCAACCCACCTCCAGATGCTGCTGGCAAGGCGAGC 1261
Qy      2495 TCAAGAGAGAGAGACAGCGCGCTTGGCAGACATACAGACACCAACCCAAAGGCGTGG 2554
Db      1262 TCAAGAGAGAGAGACAGCGCGCTTGGCAGACATACAGACACCAACCCAAAGGCGTGG 1321
Qy      2555 ACCCCAGGCGCGCGCTGCTGATCCAGCAGGTGGCACTGACGCTCCCGCTCTTCGACGT 2614
Db      1322 ACCCCAGGCGCGCGCTGCTGATCCAGCAGGTGGCACTGACGCTCCCGCTCTTCGACGT 1381
Qy      2615 CCAAGGCTCTCAGAGAAACACAGGCGCTGCTGCTCCGAGACCTTCTCTGACACCTTCTCT 2674
Db      1382 CCAAGGCTCTCAGAGAAACACAGGCGCTGCTGCTCCGAGACCTTCTCTGACACCTTCTCT 1441
Qy      2675 CCAAGGCTCTCAGAGAAACACAGGCGCTGCTGCTCCGAGACCTTCTCTGACACCTTCTCT 2734
Db      1442 CCAAGGCTCTCAGAGAAACACAGGCGCTGCTGCTCCGAGACCTTCTCTGACACCTTCTCT 1501
Qy      2735 CTGGGCTGCAAGTAAGTTACGTGAGGCTTCTCTCGGGGCTTGGAAAGTGGCAGCCATCA 2794
Db      1502 CTGGGCTGCAAGTAAGTTACGTGAGGCTTCTCTCGGGGCTTGGAAAGTGGCAGCCATCA 1561
Qy      2795 GTTGTCTTGTGTGACCCCTCGAGACAGCGCCGACAGGTGTGTGTGTGTGTGTGTGTGTGTGT 2854
Db      1562 GTTGTCTTGTGTGACCCCTCGAGACAGCGCCGACAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1621
Qy      2855 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2914
Db      1622 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1681
Qy      2915 AACCTCCCGACGACTGTATCTGAGTAAGCGGCTTAAGAACTTCTCTCTGCTTGTGTGTGTGT 2974
Db      1682 AACCTCCCGACGACTGTATCTGAGTAAGCGGCTTAAGAACTTCTCTCTGCTTGTGTGTGT 1741
Qy      2975 AGGAGTTCGGGTTGTCCAAATTCGTGAACATTCATTCATTTTAAAGGTTTCTCTCTCTCT 3034
Db      1742 AGGAGTTCGGGTTGTCCAAATTCGTGAACATTCATTCATTTTAAAGGTTTCTCTCTCTCT 1800
Qy      3035 GACGGGCGCCCAAGCGCGCGGCGGTGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3094
Db      1801 GACGGGCGCCCAAGCGCGCGGCGGTGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Qy      3095 TCCCGTGGCGCTTGTGCGGAGGTGCTTGTGGGCAACCCATCTCTGCGTTTCATTTGACGT 3154
Db      1861 TCCCGTGGCGCTTGTGCGGAGGTGCTTGTGGGCAACCCATCTCTGCGTTTCATTTGACGT 1920
Qy      3155 CGACTGTACAGAAAGGCACTCACCACAATAAACCCTTCTGAAAGCGAAGAAAAA 3213
Db      1921 CGACTGTACAGAAAGGCACTCACCACAATAAACCCTTCTGAAAGCGAAGAAAAA 1979

```

```

RESULT 7
LOCUS   BC014145          2125 bp    mRNA    linear    PRI 19-DEC-2003
DEFINITION Homo sapiens cDNA clone MGC:20641 IMAGE:4762069, complete cds.
ACCESSION BC014145.2 GI:40226303
VERSION   BC014145.2
KEYWORDS  MGC.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS   Klausner,R.D., Collins,F.S., Wagner,L.H., Grouse,L.H., Derge,J.G.,
            Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stappleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Teshitky,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullaney,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butlerfield,Y.S., Krzywinski,M.I., Skleleka,U., Small,D.E.,
            Scherker,A., Schein,J.E., Jones,S.J., and Marra,M.A.
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
TITLE      2 (bases 1 to 2125)
JOURNAL    Strausberg,R.
REFERENCE  Direct Submission
            Submitted (10-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
            On Dec 19, 2003 this sequence version replaced gi:15559567.
COMMENT    Contact: MGC help desk
            Email: cgabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland:
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nigr1.nih.gov
            Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaspi,R.,
            Maduro,O.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Scantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgonov,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information found
            through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
            Series: IRAL Plate: 30 Row: 0 Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8922628.
FEATURES   Location/Qualifiers
            1..2125
            source          "Homo sapiens"
            mol_type        "mRNA"
            db_xref="taxon:9606"
            /clone="MGC:20641 IMAGE:4762069"
            /issue_type="Brain, neuroblastoma"
            /clone_lib="NIH MGC_47"
            /lab_host="DH10B-R"
            /note="vector: pOTB7"
            15..1538
            /codon_start=1
            /product="Unknown (protein for MGC:20641)"
            /protein_id="AAH14145.1"
            /db_xref="GI:15559568"
            /translation="MATHLSEELDNEEDYSLNVRRENSSEELKAATYRLCMLYHPD
            KHRDELKSOAERLNLVHQAIEVLSDDQTRAIYIYKRGLEMGWEVERRRPAP
            IREPERLQREBERRLQOATPKOTIEVGDATDLPBYEYEDVSGSPPOEIN
            KMHISQIEAPLATDTAIIISGISTONGGGSINPALRYVTSKMGSELEFGAGDL
            OGPLTGKLFRLNLTPTRCVTNCAIOPSRGIRPGLTTLVARNLKNKYGVLYOWRGI
            OSHNLSIVYDRTKSHFTVALQDIPHSFALISYHKKODDDQTRTKSGKAGFGTV
            VAYGKERKISRISVGAIVSGVGVSLKVELEKORESAATDVLOKKQKESAVRL
            MGSVARIIEAESRMGLIIVAWGKRVNDSRSEKVKYIDVTVPLOCLVQDKLI
            ILTEAKAGIPGVDPCEVEEKILKLYQFRGLVHGVMLDSRLAIPQOSHRIDTDG"
            54..2260
            /note="DnaJ. Region: DnaJ domain. DnaJ domains (J-domains)
            are associated with hsp70 heat-shock system and it is
            thought that this domain mediates the interaction.
            DnaJ-domain is therefore part of a chaperone (protein

```

folding) system. The T-antigens are confirmed as DnaU  
containing domains from literature"  
/db\_xref="CDD:pfam00226"

## ORIGIN

Query Match 60.2%; Score 1934; DB 9; Length 2125;  
Best Local Similarity 93.1%; Pred. No. 0;  
Matches 2100; Conservative 0; Mismatches 0; Indels 156; Gaps 1;

QY 1 GAAAGGTTGCGAAGTGGCGACGCGCTTTGAGCGAGAGAGCTGACATGAAGCTATT 60  
DB 1 GAAAGGTTGCGAAGTGGCGACGCGCTTTGAGCGAGAGAGCTGACATGAAGCTATT 60  
QY 61 ACTGTTGCTGAAGCTGCGAGAGAGCGCTTTGAGAGAGCTGAGAGAGCTGAGAGAG 120  
DB 61 ACTGTTGCTGAAGCTGCGAGAGAGCGCTTTGAGAGAGCTGAGAGAGCTGAGAGAG 120  
QY 121 GGCTCTGATGCTCTACCATCCAGACAGAGAGCCAGAGCTCAAGTCAAGGCGG 180  
DB 121 GGCTCTGATGCTCTACCATCCAGACAGAGAGCCAGAGCTCAAGTCAAGGCGG 180  
QY 181 AAGCAGTGTAAACCTTGTTCACAGGCTTATGAAGTCTTAGTACCCCAACAGGG 240  
DB 181 AAGCAGTGTAAACCTTGTTCACAGGCTTATGAAGTCTTAGTACCCCAACAGGG 240  
QY 241 CCATCTATGATATATATGAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 241 CCATCTATGATATATATGAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 300  
QY 301 GAG 360  
DB 301 GAG 360  
QY 361 AG 420  
DB 361 AG 420  
QY 421 CCGAGCTTTTGTATGCGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 421 CCGAGCTTTTGTATGCGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 481 AGATTGAATTAATTAATTAATGACATATCCAGTCCATTGAGAGAGAGAGAGAGAG 540  
DB 481 AGATTGAATTAATTAATTAATGACATATCCAGTCCATTGAGAGAGAGAGAGAGAG 540  
QY 541 ACACAGCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 541 ACACAGCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 601 ACTTTGGCTCAGACGAGTAACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 ACTTTGGCTCAGACGAGTAACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 GAGAGCTAAG 720  
DB 661 GAGAGCTAAG 720  
QY 721 TTGTGACAAACAACTGTGCTGAGAGTTCATCCCGTGAATCCGAGCCGCTGACCA 780  
DB 721 TTGTGACAAACAACTGTGCTGAGAGTTCATCCCGTGAATCCGAGCCGCTGACCA 780  
QY 781 CTGTCTTATGCTGAG 840  
DB 781 CTGTCTTATGCTGAG 840  
QY 841 TCCAGTCAAG 900  
DB 841 TCCAGTCAAG 900  
QY 901 CCTGCAAG 960  
DB 901 CCTGCAAG 960

QY 961 ATGACGATCAGAGCTCGTGTGAAAGAGATCCCTCAAGAGAGAGCTTTGGAGAGAGTGG 1020  
DB 961 ATGACGATCAGAGCTCGTGTGAAAGAGATCCCTCAAGAGAGAGCTTTGGAGAGAGTGG 1020  
QY 1021 AGTACGAGAGCTGAGAGAGAGATCTCCAGGACAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1021 AGTACGAGAGCTGAGAGAGAGATCTCCAGGACAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 GAGTTCACAG 1140  
DB 1081 GAGTTCACAG 1140  
QY 1141 TCCCTATTCTTGAAG 1200  
DB 1141 TCCCTATTCTTGAAG 1200  
QY 1201 CTCTAGTGTCTACTTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 CTCTAGTGTCTACTTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 AAG 1320  
DB 1261 AAG 1320  
QY 1321 AAG 1380  
DB 1321 AAG 1380  
QY 1381 AAG 1440  
DB 1381 AAG 1440  
QY 1441 AAG 1500  
DB 1441 AAG 1500  
QY 1501 TGAAG 1560  
DB 1501 TGAAG 1560  
QY 1561 ACCGCTGTGTGAG 1620  
DB 1561 ACCGCTGTGTGAG 1620  
QY 1621 ATCAG 1680  
DB 1621 ATCAG 1680  
QY 1681 ATCAG 1740  
DB 1681 ATCAG 1740  
QY 1741 GAGAGCTTACAAATTTGAG 1800  
DB 1741 GAGAGCTTACAAATTTGAG 1800  
QY 1801 TTAATTAAG 1860  
DB 1801 TTAATTAAG 1860  
QY 1861 TTAATTAAG 1920  
DB 1861 TTAATTAAG 1920  
QY 1921 CATGAGCTGTTTCCAAAGAGATGAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
DB 1921 CATGAGCTGTTTCCAAAGAGATGAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
QY 1981 TGTCCGAG 2040  
DB 1981 TGTCCGAG 2040  
QY 2041 GGTGTGTGTGAG 2100  
DB 2041 GGTGTGTGTGAG 2100



Db 1885 GGTGTCTGTGAGACACCTCCCTCACACAGAGAGGTGAGCCATTCTGACATCCACGC 1944  
QY 2101 TGAACATGGGAAACAACCTGAAGACGAGAGCCTCCCGTCAAGGAGCCTCTGCTG 2160  
Db 1945 TGAACATGGGAAACAACCTGAAGACGAGAGCCTCCCGTCAAGGAGCCTCTGCTG 2004  
QY 2161 CTGGCTCCCATGACCACTCTCTCTGCTGAATAATTACTGTGAATCTGAGCAGATT 2220  
Db 2005 CTGGCTCCCATGACCACTCTCTCTGCTGAATAATTACTGTGAATCTGAGCAGATT 2064  
QY 2221 GCGGGTTATTAACATGCTTTTATCTGAGACAAA 2256  
Db 2065 GCGGGTTATTAACATGCTTTTATCTGAGACAAA 2100  
RESULT 8  
AX748271 2454 bp mRNA linear PAT 20-JUN-2003  
LOCUS Sequence 1796 from Patent EP1308459.  
DEFINITION AX748271  
ACCESSION AX748271 GI:32132659  
VERSION AX748271.1 GI:32132659  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
Iosagai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,  
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,  
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and  
Masuko,Y.  
Full-length cDNA sequences  
Patent: EP 1308459-A 1796 07-MAY-2003:  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)  
LOCATION/Qualifiers  
FEATURES  
source 1..2454  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 57.6%; Score 1851.6; DB 6; Length 2454;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1338 GTCCGGCTGATGCAAGAACTGTCCGAAGATTAATTGAGCAGAAAGTCCAGAATGGGC 1397  
Db 585 GTCCGGCTGATGCAAGAACTGTCCGAAGATTAATTGAGCAGAAAGTCCAGAATGGGC 644  
QY 1398 CTGATCATGCTCAATGCTGTGTACGGGAAGTTTGTCAATGACAAAGCAGAGAGACGAG 1457  
Db 645 CTGATCATGCTCAATGCTGTGTACGGGAAGTTTGTCAATGACAAAGCAGAGAGACGAG 704  
QY 1458 AAGGTGAAGGTGATTGACGTGACTGTGCCCCCTGACAGTCTGTGTAAGACTCGAAGCTC 1517  
Db 705 AAGGTGAAGGTGATTGACGTGACTGTGCCCCCTGACAGTCTGTGTAAGACTCGAAGCTC 764  
QY 1518 ATCTTCAGGAGGCTCTCAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGGAA 1577  
Db 765 ATCTTCAGGAGGCTCTCAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGGAA 824  
QY 1578 GAGAAAGACCTGAAGTGTCTATCAGTTCCGGGGGCTCTGATCAGGTGATGTGCTG 1637  
Db 825 GAGAAAGACCTGAAGTGTCTATCAGTTCCGGGGGCTCTGATCAGGTGATGTGCTG 884  
QY 1638 GACAGTGAAGCCTCTCCGATATCAAAAGCAGTCCACAGATCGATACAGATGATTAAC 1697  
Db 885 GACAGTGAAGCCTCTCCGATATCAAAAGCAGTCCACAGATCGATACAGATGATTAAC 944  
QY 1698 GCCAAGAACCAAGATTTTAAAGGCCGCAAAAATCTTTTCTGGAGATCAAAATTTG 1757

Db 945 GCCAAGAACCAAGATTTTAAAGGCCGCAAAAATCTTTTCTGGAGATCAAAATTTG 1004  
QY 1758 GAAATGAAGAAACCCAGACATGATGTTTATTTATTTATTTATTTATTAAGGTGT 1817  
Db 1005 GAAATGAAGAAACCCAGACATGATGTTTATTTATTTATTTATTTATTAAGGTGT 1064  
QY 1818 ACCATTATCAATTTATGTAAGGAGATGACAGACACCCAGCTTTTGAGGGTCTGGGGT 1877  
Db 1065 ACCATTATCAATTTATGTAAGGAGATGACAGACACCCAGCTTTTGAGGGTCTGGGGT 1124  
QY 1878 AGGACTGAGGCGCCCACTGGGAACCAAGCTGCAAGCTGGGCCATGTGCTTTTCCCA 1937  
Db 1125 AGGACTGAGGCGCCCACTGGGAACCAAGCTGCAAGCTGGGCCATGTGCTTTTCCCA 1184  
QY 1938 GGATCAGTCTCTGAGGGAAGGGCTGTGGCCCTGACTCCGCTGTGTCCGAGACACG 1997  
Db 1185 GGATCAGTCTCTGAGGGAAGGGCTGTGGCCCTGACTCCGCTGTGTCCGAGACACG 1244  
QY 1998 CTGACCGAGCCCGCCCTGTAGTCTTGTGCTGGGCTGAGAGTGTCTGAGACACC 2057  
Db 1245 CTGACCGAGCCCGCCCTGTAGTCTTGTGCTGGGCTGAGAGTGTCTGAGACACC 1304  
QY 2058 CTGCTCTCAACAAGAGGCTGAGCCATTCTGCACTCCAGCTGAACATGGGAAACAAC 2117  
Db 1305 CTGCTCTCAACAAGAGGCTGAGCCATTCTGCACTCCAGCTGAACATGGGAAACAAC 1364  
QY 2118 CTGAAGAAGAGGAGGCTCCCGGTCAAGGAGCCTGTGCTGTGCTGCTCCATGACCA 2177  
Db 1365 CTGAAGAAGAGGAGGCTCCCGGTCAAGGAGCCTGTGCTGTGCTGCTCCATGACCA 1424  
QY 2178 CCTCTCTCTGTAATATTTACTGTGATCTGAGCAGATTTGGGGTTTAAACCTG 2237  
Db 1425 CCTCTCTCTGTAATATTTACTGTGATCTGAGCAGATTTGGGGTTTAAACCTG 1484  
QY 2238 CTTTTTATCTGGAACAACCGGTTTGAATAATTAAGTCTTTTCCCATCCCAAG 2297  
Db 1485 CTTTTTATCTGGAACAACCGGTTTGAATAATTAAGTCTTTTCCCATCCCAAG 1544  
QY 2298 CTGCTCAAGTCAATTCACAGGCCCCCTCGGCTTGGGACAGGGTAGTAACTCCGATCC 2357  
Db 1545 CTGCTCAAGTCAATTCACAGGCCCCCTCGGCTTGGGACAGGGTAGTAACTCCGATCC 1604  
QY 2358 CAGGGCTAGCCCTGACACAGGTGTCTCCCGTATCCCGGTGGGAAACGCCCTGCACC 2417  
Db 1605 CAGGGCTAGCCCTGACACAGGTGTCTCCCGTATCCCGGTGGGAAACGCCCTGCACC 1664  
QY 2418 AGCGGCTTAGCTGGCTGTGTCTCTCAAGCCTGCACACCCACTCTCAAGTGCAG 2477  
Db 1665 AGCGGCTTAGCTGGCTGTGTCTCTCAAGCCTGCACACCCACTCTCAAGTGCAG 1724  
QY 2478 TGTGTGGCAAGGGCAGCTCAAGAGACAGGACAGAGCGCTTGGCAAGACATGACACAC 2537  
Db 1725 TGTGTGGCAAGGGCAGCTCAAGAGACAGGACAGAGCGCTTGGCAAGACATGACACAC 1784  
QY 2538 CCAACCCAAAGGCTGAGACCCAGGCGCCGCTGTGTAACCAAGAGTGGACTGCACG 2597  
Db 1785 CCAACCCAAAGGCTGAGACCCAGGCGCCGCTGTGTAACCAAGAGTGGACTGCACG 1844  
QY 2598 TCCCGCTCTGCAAGTCCAGGCTCTCAAGAGACCAAGGCGCTGTGCTCCGAGCT 2657  
Db 1845 TCCCGCTCTGCAAGTCCAGGCTCTCAAGAGACCAAGGCGCTGTGCTCCGAGCT 1904  
QY 2658 TCCCTCAGACCTCTCTCCAGCTGACCTCACTTGGAGTGCAGATGACAGGAGCTAGGACC 2717  
Db 1905 TCCCTCAGACCTCTCTCCAGCTGACCTCACTTGGAGTGCAGATGACAGGAGCTAGGACC 1964  
QY 2718 CCTTCAGGCTCTGACCTCTGAGCTCAAGTAAAGTTACGTGAGCCTGTCTCTGGGGCCT 2777  
Db 1965 CCTTCAGGCTCTGAGCTCTGAGCTCAAGTAAAGTTACGTGAGCCTGTCTCTGGGGCCT 2024  
QY 2778 GGAAGTGGAGCAATCAAGTGTCTTTGCTGACCCCTCGAAGCAAGCGCCGCAAGGTGT 2837  
Db 2025 GGAAGTGGAGCAATCAAGTGTCTTTGCTGACCCCTCGAAGCAAGCGCCGCAAGGTGT 2084

2838 GGTGAGACAGCTGGCGGGGGGCCCCAGCTGGCCGGCTCCAGCCCACCACAGCT 2897  
2085 GGCTGAGACAGCTGGCGGGGGGCCCCAGCTGGCCGGCTCCAGCCCACCACAGCT 2144  
2898 GTTGTGTAAGTCAGGCCAAGCTCCCGACAGCTGATATCTGATAGTAAGGTAAGCTCC 2957  
2145 GTTGTGTAAGTCAGGCCAAGCTCCCGACAGCTGATATCTGATAGTAAGGTAAGCTCC 2204  
2958 TTCTCTGTGTTTGTAAAAAGCATTTGGGTTGTCCATTCTGTAACTTCAATCTCATTTT 3017  
2205 TTCTCTGTGTTTGTAAAAAGCATTTGGGTTGTCCATTCTGTAACTTCAATCTCATTTT 2264  
3018 TTAAGAAAGTTTCTGTAGAGGCCCGACAGGCCCGGAGCGCGGTGAGCGCTCGTTGCAATG 3077  
2265 TT-AAAAAGTTTCTGTAGAGGCCCGACAGGCCCGGAGCGCGGTGAGCGCTCGTTGCAATG 2323  
3078 AGCTGTGGCGCCCGGGCTTCCGCTGGCGCTGTGCGCAGAGTCTTCTGGGCAACCATCTCTC 3137  
2224 AGCTGTGGCGCCCGGGCTTCCGCTGGCGCTGTGCGCAGAGTCTTCTGGGCAACCATCTCTC 2383  
3138 TGCCTTCAATTTGACAGTCTGATGACAGAGGCACTCAACAATTAACCTTTCTGTAA 3197  
2384 TGCCTTCAATTTGACAGTCTGATGACAGAGGCACTCAACAATTAACCTTTCTGTAA 2443  
3198 GCAGAAAAA 3207  
2444 GCAAAAAA 2453

RESULT 9  
AK093709 2454 bp mRNA linear PRI 06-SEP-2003  
LOCUS  
DEFINITION Homo sapiens cDNA FLJ36390 f18, clone THYMJ2008990.  
ACCESSION AK093709  
VERSION AK093709.1 GI:21752634  
KEYWORDS oligo capping; f18 (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1  
Oshima,A., Takahashi,Fuji,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Muraashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isegai,T.  
MEDO human cDNA sequencing project  
JOURNAL  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 2454)  
AUTHORS Isegai,T. and Yamamoto,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isegai, FLJ Project (HRI Team); 2-6-7  
Kasue-Kamatari, Kibatazu, Chiba 292-0812, Japan  
E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source  
1..2454  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="THYMJ2008990"  
/tissue\_type="thymus"

ORIGIN  
/clone\_11b="THYMJ2"  
/note="Cloning vector: pME18SFL3"  
Query Match 57.6%; Score 1851.6; DB 9; Length 2454;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
1338 GTCCGCTGATGACAGAAATCTCTCCGAAGATTAATGAGGCAAGAGTCCAGAAATGGC 1397  
585 GTCCGCTGATGACAGAAATCTCTCCGAAGATTAATGAGGCAAGAGTCCAGAAATGGC 644  
1398 CTGATCAATCGTCAATGCTGCTGTAAGGAAATTTGTCAATGACAGAGCAGAAAGAGAG 1457  
645 CTGATCAATCGTCAATGCTGCTGTAAGGAAATTTGTCAATGACAGAGCAGAAAGAGAG 704  
1458 AAGGTGAAGGTGATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1517  
705 AAGGTGAAGGTGATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 764  
1518 ATCTCAAGAGGCTCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577  
765 ATCTCAAGAGGCTCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824  
1578 GAGAAAGACCTGAAGGCTCTATACGTTCCGGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1637  
825 GAGAAAGACCTGAAGGCTCTATACGTTCCGGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 884  
1638 GACAGTAGAGGCTCTCCGATATCCAAAGCAGTCCCAAGATTCATGATGATGATTAACCT 1697  
885 GACAGTAGAGGCTCTCCGATATCCAAAGCAGTCCCAAGATTCATGATGATGATTAACCT 944  
1698 GCCAAGACCAAGATTTTAAAGGCGCCCAAAATCTTTTCTGCGAGTCTCAAAATTTG 1757  
945 GCCAAGACCAAGATTTTAAAGGCGCCCAAAATCTTTTCTGCGAGTCTCAAAATTTG 1004  
1758 GAAATGAAAAAACCCAGACATCAGATGTTTATTTATTTATTTATTTATTTATTTATTTATTT 1817  
1005 GAAATGAAAAAACCCAGACATCAGATGTTTATTTATTTATTTATTTATTTATTTATTTATTT 1064  
1818 ACCATTATCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1877  
1065 ACCATTATCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1124  
1878 AGAGCTGAGGAGCCCACTGAGGAAACAGACATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1937  
1125 AGAGCTGAGGAGCCCACTGAGGAAACAGACATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1184  
1938 GAGTCAATTTCTGAGAGGAAAGGCTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1997  
1185 GAGTCAATTTCTGAGAGGAAAGGCTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1244  
1998 CTGACCGCAGCGCGCGCGCTGAGTTCTTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 2057  
1245 CTGACCGCAGCGCGCGCGCTGAGTTCTTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1304  
2058 CTGCTCTACCAAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2117  
1305 CTGCTCTACCAAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1364  
2118 CTGAAAAGCAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 2177  
1365 CTGAAAAGCAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1424  
2178 CTGCTCTCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2237  
1425 CTGCTCTCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1484  
2238 CTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2297  
1485 CTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1544  
2298 CTGCTCAAGTATTTCCAGCGGCGCGCTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 2357

Db 1545 CTCTCAAGTCATTCACCGGCCCCCTCGGCTTGAGGACAGGTAAGTAATCCCATCC 1604  
 2358 CAGGGGCTAGCCCTGACAGAGTGGCTTCCGTAATCCCGGTGGAAAAAGCCCTGACAC 2417  
 Db 1605 CAGGGGCTAGCCCTGACAGAGTGGCTTCCGTAATCCCGGTGGAAAAAGCCCTGACAC 1664  
 QY 2418 AGCGGGCTTAGAGTGGCTTGTCTCCCTCAACCGCTGACACCACTTCAGAGTGCAG 2477  
 Db 1665 AGCGGGCTTAGAGTGGCTTGTCTCCCTCAACCGCTGACACCACTTCAGAGTGCAG 1724  
 QY 2478 TGTCTGGGCAAGGGAGCTCAAGAGACAGACAGGCGCTTGGAGACATCAGACACAC 2537  
 Db 1725 TGTCTGGGCAAGGGAGCTCAAGAGACAGACAGGCGCTTGGAGACATCAGACACAC 1784  
 QY 2538 CCAACCCCAAGGGGTGAGACCCAGGCGCGGCTGTACCCAGAGGTGGCACTGAGC 2597  
 Db 1785 CCAACCCCAAGGGGTGAGACCCAGGCGCGGCTGTACCCAGAGGTGGCACTGAGC 1844  
 QY 2598 TCCCGGCTCTGAGAGTCCAGCGTCTCAAGAGAACACAGGGCTGTGCTCCGAGCCT 2657  
 Db 1845 TCCCGGCTCTGAGAGTCCAGCGTCTCAAGAGAACACAGGGCTGTGCTCCGAGCCT 1904  
 QY 2658 TCCCTTCAAGACCTTCTCTCCAGTGCACCTTGGAGATGACAGCGGAGCTAGGACC 2717  
 Db 1905 TCCCTTCAAGACCTTCTCTCCAGTGCACCTTGGAGATGACAGCGGAGCTAGGACC 1964  
 QY 2718 CCTTCAAGCGGCTGAGACCTGAGTGAATTAAGTGAAGGCTGTCTGAGGCGCT 2777  
 Db 1965 CCTTCAAGCGGCTGAGACCTGAGTGAATTAAGTGAAGGCTGTCTGAGGCGCT 2024  
 QY 2778 GGAAGTGGCAGCAGCATAGTGTCTTGTGCTGACCTCCAGAGACGCGCCAGAGTGT 2837  
 Db 2025 GGAAGTGGCAGCAGCATAGTGTCTTGTGCTGACCTCCAGAGACGCGCCAGAGTGT 2084  
 QY 2838 GGTGTGAGACAGCTGGCGCGGCGGCGCCAGAGCTGGCGCGCTCCAGCCCAAGAGCT 2897  
 Db 2085 GGTGTGAGACAGCTGGCGCGGCGGCGCCAGAGCTGGCGCGCTCCAGCCCAAGAGCT 2144  
 QY 2898 GTTGTGAGACAGCTGGCGCGGCGGCGCCAGAGCTGGTAATGTAAGGCTTAAGAGCTCC 2957  
 Db 2145 GTTGTGAGACAGCTGGCGCGGCGGCGCCAGAGCTGGTAATGTAAGGCTTAAGAGCTCC 2204  
 QY 2958 TTCTCTGCTTTTGAAGAAGAGTTCGGGTTGTCATTTCTGTAATCTCAATCTCAATTT 3017  
 Db 2205 TTCTCTGCTTTTGAAGAAGAGTTCGGGTTGTCATTTCTGTAATCTCAATCTCAATTT 2264  
 QY 3018 TTTAAAAAGGTTTCTGACGCGCCCAAGGCGCGGCTGAGCGCTGTGTCATG 3077  
 Db 2265 TTTAAAAAGGTTTCTGACGCGCCCAAGGCGCGGCTGAGCGCTGTGTCATG 2323  
 QY 3078 AGGCTGGGCGCGGCGGCTTCCGAGCGCTGTGCGAGGCTTTCGGGAGACCAATCTTC 3137  
 Db 2324 AGGCTGGGCGCGGCGGCTTCCGAGCGCTGTGCGAGGCTTTCGGGAGACCAATCTTC 2383  
 QY 3138 TGGCTTTCAATTTGACGCTGACGCTGACAGAGGAGCACTCACCAATTAACCTTTCTGAAA 3197  
 Db 2384 TGGCTTTCAATTTGACGCTGACGCTGACAGAGGAGCACTCACCAATTAACCTTTCTGAAA 2443  
 QY 3198 GCAGAAAAA 3207  
 Db 2444 GCAGAAAAA 2453

RESULT 10  
 AK095386  
 LOCUS Homo sapiens cDNA FLJ38067 fis, clone CTONG2015345, weakly similar  
 DEFINITION to DnaJ protein.  
 ACCESSION AK095386  
 VERSION AK095386.1 GI:21754632  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kikuchi, H., Kanda, K., Matsuda, H., Matsuda, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Maehuo, Y., Nagai, K. and Isogai, T.  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3012)  
 AUTHORS Isogai, T. and Yamamoto, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
 FEATURES  
 source  
 1. .3012  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CTONG2015345"  
 /issue\_type="tongue, tumor tissue"  
 /clone\_lib="CTONG2"  
 /note="Cloning vector: pME185FL3"  
 ORIGIN  
 Query Match 57.4%; Score 1845.2; DB 9; Length 3012;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1858; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
 QY 1338 GTCCGCGTATCAGAGATCTGTCCAGAGATTAATTGAGCAGAAAGTCCAGAAATGGGC 1397  
 Db 1152 GTCCGCGTATCAGAGATCTGTCCAGAGATTAATTGAGCAGAAAGTCCAGAAATGGGC 1211  
 QY 1398 CTCATCATGTCATATGCTGTGACGGGAAGTTTGTCAATGACAGAGCAGAAAGCGAG 1457  
 Db 1212 CTCATCATGTCATATGCTGTGACGGGAAGTTTGTCAATGACAGAGCAGAAAGCGAG 1271  
 QY 1458 AAGGTGAAGTATGACGTCGTCGCGCGCTGCGAGTGGTGAAGACTCGAAGCTC 1517  
 Db 1272 AAGGTGAAGTATGACGTCGTCGCGCGCTGCGAGTGGTGAAGACTCGAAGCTC 1331  
 QY 1518 ATCTTCACGAGAGGCTCCAGAGCTGGGCTGCTGCTTTATGACCCGCTGTGGGGGAA 1577  
 Db 1332 ATCTTCACGAGAGGCTCCAGAGCTGGGCTGCTGCTTTATGACCCGCTGTGGGGGAA 1391  
 QY 1578 GAGAGAACCTGAAGTGTCTATCATGTTCCGGGGGCTCTGCATCAGGTGATGCTG 1637  
 Db 1392 GAGAGAACCTGAAGTGTCTATCATGTTCCGGGGGCTCTGCATCAGGTGATGCTG 1451  
 QY 1638 GACAGTGAAGCCCTCCGATTAACCAAGAGTCCCAAGATGATGATGATGATTAAC 1697  
 Db 1452 GACAGTGAAGCCCTCCGATTAACCAAGAGTCCCAAGATGATGATGATGATTAAC 1511  
 QY 1698 GCAGAGAACCATTTTAAAGGCGCAAAATCTTTCTCGGAGTCTCAAAATTTG 1757  
 Db 1512 GCAGAGAACCATTTTAAAGGCGCAAAATCTTTCTCGGAGTCTCAAAATTTG 1571  
 QY 1758 GAAATGAAAAAACCAGACATGATGTTTATTTATTTATTTATTTATGAGGTGCT 1817

Db	1572	GAAAAAGAAAAA	CCAGACATCAGATGTTTTTTATTTATATATATATATATAGAAAGTGGT	1631
QY	1818	ACCATTTATCATTTATATGTGAAGGACATGCAGACACCCAGCTTTTGAAGGGTGTGGGGGT	1877	
Db	1632	ACCATTTATCATTTATATGTGAAGGGAATGTGACAGACCCCGAGCTTTTGAAGGGTGTGGGGGT	1691	
QY	1878	AGGATGAGGAGGAGCCCACTGGGAATCCAGACTGAGAGCTTGGGCCATTGGCTGTGTTTCCAA	1933	
Db	1692	AGGACTGAGGGCAGCCCCCACTGGGAACAGACTGAGAGCTTGGGCCATTGGCTGTGTTTCCAA	1751	
QY	1938	GGATCAGTTCCGAGAGGGAAGGGCTCTGGGCCCTGACTCCGCTGTGTCCCGAGACACGCTG	1997	
Db	1752	GGATCAGTTCCGAGAGGGAAGGGCTCTGGGCCCTGACTCCGCTGTGTCCCGAGACACGCTG	1811	
QY	1998	CTGACCGAGAGCCCGCCGCTGTAGTTCCTTGGCTGGGATCTGGAGGTGTCTGTGAGCAC	2057	
Db	1812	CTGACCGAGAGCCCGCCGCTGTAGTTCCTTGGCTGGGATCTGGAGGTGTCTGTGAGCAC	1874	
QY	2058	CTGCGCCTCACCAAGAGAGCTGAGAGCATTCCTGACGTCCAGCTGAACATGGGAAACAAC	2117	
Db	1872	CTGCGCCTCACCAAGAGAGCTGAGAGCATTCCTGACGTCCAGCTGAACATGGGAAACAAC	1931	
QY	2118	CTGAAAAAGCAGGACGGGCTCCCGGCTCAGGAGGCTCTGTGCTGGCTGGCTCCCAATGACA	2177	
Db	1932	CTGAAAAAGCAGGACGGGCTCCCGGCTCAGGAGGCTCTGTGCTGGCTGGCTCCCAATGACA	1991	
QY	2178	CCTCCTCTGCTGAATAATATATCTGCTTGAATCTGAGACAGATTCGGGGTTTATAAACTG	2237	
Db	1992	CCTCCTCTGCTGAATAATATATCTGCTTGAATCTGAGACAGATTCGGGGTTTATAAACTG	2051	
QY	2238	CTTTTATATCTGAGAACAAACGGGTTTGGAAATTAGTGTCTTTTTCCTTCCCACTCCAGAG	2297	
Db	2052	CTTTTATATCTGAGAACAAACGGGTTTGGAAATTAGTGTCTTTTTCCTTCCCACTCCAGAG	2111	
QY	2298	CTGCTCAAGTCAATTCACACGGGCCCCCTGGGCTTGGGACAGGGGTAGTGAATCMTCCGATCC	2357	
Db	2112	CTGCTCAAGTCAATTCACACGGGCCCCCTGGGCTTGGGACAGGGGTAGTGAATCMTCCGATCC	2171	
QY	2358	CAGGGCTTAGCCCTTGACACAGGTGGCTTCCCGTATCCCGTGTGGAAAAAGCCCTTGCCAC	2417	
Db	2172	CAGGGCTTAGCCCTTGACACAGGTGGCTTCCCGTATCCCGTGTGGAAAAAGCCCTTGCCAC	2231	
QY	2418	AGCGGGCTTGAAGCTGGCTGTGTCCCTCCACCGCTTGACACACCACTCCAGAGTGCAG	2477	
Db	2232	AGCGGGCTTGAAGCTGGCTGTGTCCCTCCACCGCTTGACACACCACTCCAGAGTGCAG	2291	
QY	2478	TGCTGGGCAAGGGGAGCTCAAGAGGACAGAACAGGGGCTTGGGACATGCACACAC	2537	
Db	2232	TGCTGGGCAAGGGGAGCTCAAGAGGACAGAACAGGGGCTTGGGACATGCACACAC	2351	
QY	2538	CCAACCCCAAAGGCGTGGACCCCAAGGCCCGGCGCGGTGTATCCAGAGGTGGCATCTGCAGC	2597	
Db	2352	CCAACCCCAAAGGCGTGGACCCCAAGGCCCGGCGCGGTGTATCCAGAGGTGGCATCTGCAGC	2411	
QY	2598	TCCCGGCTCTGCAAGTCCAGGCTCTCTACAGGAACACCAAGGCTGTGCTCCGAGGCT	2657	
Db	2412	TCCCGGCTCTGCAAGTCCAGGCTCTCTACAGGAACACCAAGGCTGTGCTCCGAGGCT	2471	
QY	2658	TCCCTTCAAGACCTTCTCTCAAGTGGCCACTTGGGATGAGAAATGAGGGGAGCTTAGAGAC	2717	
Db	2472	TCCCTTCAAGACCTTCTCTCAAGTGGCCACTTGGGATGAGAAATGAGGGGAGCTTAGAGAC	2531	
QY	2718	CCCTTCAAGGAGCTTGAGACTTCGGCTGCAATTAAGTTACGTGAGGCTGTCTCTCGGGGCT	2777	
Db	2532	CCCTTCAAGGAGCTTGAGACTTCGGCTGCAATTAAGTTACGTGAGGCTGTCTCTCGGGGCT	2591	
QY	2778	GGAAGTGGCAGCATCAGTGTCTTGTCTGACCCCTCGAGCAAGAGCGCGCACAGTGGT	2837	
Db	2592	GGAAGTGGCAGCATCAGTGTCTTGTCTGACCCCTCGAGCAAGAGCGCGCACAGTGGT	2651	
QY	2838	GGCTGAGACAGCTGGCGGGGGGGGCCAACCTGGCGCGGCTCAGGCCCAACCAAGCT	2897	
Db	2652	GGCTGAGACAGCTGGCGGGGGGGGCCAACCTGGCGCGGCTCAGGCCCAACCAAGCT	2711	

QY	2898	GTTCCTGAAGTCAGAGCCAACTCTCCAGCACTGGTACTCTAGTAACGGCTAAGAACTCC	2957
Db	2712	GTTCCTGAAGTCAGAGCCCTCCTCCAGCACTGGTACTCTAGTAACGGCTAAGAACTCC	2771
QY	2958	TTCCCTGTGTTTGAAGAAGAGTTCCGGGTTGTCCAAATTCCTGTAACTATTCCTCAATTT	3017
Db	2772	TTCCCTGTGTTTGAAGAAGAGTTCCGGGTTGTCCAAATTCCTGTAACTATTCCTCAATTT	2831
QY	3018	TTAAAAAAGGTTTCTCTGAAGGCCCCAGCCCGAGCGCGGTGAGGTGTGTGTGATG	3077
Db	2832	TT-AAAAAAGGTTTCTCTGAAGGCCCCAGCCCGAGCGCGGTGAGGTGTGTGTGATG	2890
QY	3078	AGCCTGAGCCGCCGGGCTTCCTCCGTGCGGCTCTGCGCGAGGTCTTCTGGAGCACCATCTC	3137
Db	2891	AGCCTGAGCCGCCGGGCTTCCTCCGTGCGGCTCTGCGCGAGGTCTTCTGGAGCACCATCTC	2950
QY	3138	TGCGTTTCATTTCGAGTCGACTGTACAGAAAGCACTGACCAATAAACCTTTCCTGAAA	3197
Db	2951	TGCGTTTCATTTCGAGTCGACTGTACAGAAAGCACTGACCAATAAACCTTTCCTGAAA	3010
QY	3198	GC 3199	
Db	3011	GC 3012	

RESULT 11

HS126A5

DEFINITION

Human DNA sequence from clone Rpl-126A5 on chromosome 1p36.21-36.33 Contains three novel genes (one with DnaJ domains), the gene for KIA0469 and the HKR3 gene for GLI-Kruppel family member HKR3. Contains ESTs, STS, GSSs, three CpG islands, genomic marker DIS2731 and ta and ca repeat polymorphisms, complete sequence.

AL031447

AL031447.4 GI:4826431

HTG: ca repeat polymorphism; DIS2731; DnaJ; GLI-Kruppel; HKR3; KIA0469; ta repeat polymorphism.

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 112158)

Wray, P.

Direct Submission

Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On May 13, 1999 this sequence version replaced gi:4581290.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMBASE; Information on the WORMBASE database can be found at

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Swt, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at









Db	102063	CCGACC	CGGG	GTGAG	GTGTGT	GTGATG	ATAGCT	CGGG	CCCC	CGGG	CTTCC	CGCG	CTCT	102122	
Qy	3109	GGCG	GAGT	GTCTT	CTTG	GGGAC	CCCAT	CTCTG	CGCTT	CAATT	GCAGT	CGACT	GTAC	GAG	3168
Db	102123	GGCG	GAGGT	CTTCT	TG	GGGAC	CCCAT	CTCTG	CGCTT	CAATT	GCAGT	CGACT	GTAC	GAG	102182
Qy	3169	GCAC	TCC	ACCA	CAAT	TAAC	CTTCT	CTCG	AAAG	CGA	3202				
Db	102183	GCAC	TCC	ACCA	CAAT	TAAC	CTTCT	CTCG	AAAG	CGA	102216				
RESULT 13															
AX834019/c		AX834019		1143	from Patent	EP1347046.									
LOCUS		AX834019		2052	bp	DNA									
DEFINITION		Sequence													
ACCESSION		AX834019													
VERSION		AX834019.1		GI:39920154											
KEYWORDS															
SOURCE															
ORGANISM		Homo sapiens (human)													
REFERENCE		Homo sapiens													
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.													
TITLE		1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagaharti, K. and Masuko, Y.													
JOURNAL		Full-length cDNA sequences													
FEATURES		Patent: EP 1347046-A 1143 24-SEP-2003; Research Association for Biotechnology (JP)													
source		location/Qualifiers													
ORIGIN		1..2052													
		/organism="Homo sapiens"													
		/mol_type="unassigned DNA"													
		/db_xref="taxon:9606"													
Query_Match															
Best Local Similarity		43.5%		Score 1397.4;	DB 6;	Length 2052;									
Matches 1412;		Conservative	0;	Mismatches	6;	Indels	1;	Gaps	1;						
Qy	1784	GT	TTTT	AT	TTT	AT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	1843
Db	2052	GT	TTT	AT	TTT	AT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	1993
Qy	1844	TG	CA	GA	CA	CC	CA	GC	TTT	TG	AG	GT	CT	GG	1903
Db	1992	TG	CA	GA	CA	CC	CA	GC	TTT	TG	AG	GT	CT	GG	1933
Qy	1904	CAG	ACT	GA	CA	GC	CT	TG	CC	CA	TG	AG	GT	CT	1963
Db	1932	CAG	ACT	GA	CA	GC	CT	TG	CC	CA	TG	AG	GT	CT	1873
Qy	1964	TG	GC	CT	GA	CT	CG	CT	GT	CT	CC	GA	CA	GC	2023
Db	1872	TG	GC	CT	GA	CT	CG	CT	GT	CT	CC	GA	CA	GC	1813
Qy	2024	TCT	TG	CG	CT	GA	GT	CT	GT	GA	CA	CC	CT	CA	2083
Db	1812	TCT	TG	CG	CT	GA	GT	CT	GT	GA	CA	CC	CT	CA	1753
Qy	2084	ACT	TG	CA	GT	CA	GC	CT	GA	CA	CC	CT			

Oy	2264	GGAAATTATGTCCTTTTTCCTCCCACTCCAGAGCTGCTCAAGTCATTCCACCGGCCCC	2323
Db	1572	GGAAATTATGTCCTTTTTCCTCCCACTCCAGAGCTGCTCAAAATTCACCGGCCCC	1513
Oy	2324	TCGGCTTGGGACAGGTCAGTGTAACTCCCGATCCAGGGGCTTACGCCGACACAGTGGC	2383
Db	1512	TCGGCTTGGGACAGGTCAGTGTAACTCCCGATCCAGGGGCTTACGCCCTGACACAGTGGC	1453
Oy	2384	TTCCCGTATCCCGGTGGGAAAAACGCCCTCCGACACAGCGGGCTTGAAGCTGACCTGTGTC	2443
Db	1452	TTCCCGTATCCCGGTGGGAAAAACGCCCTCCGACACAGCGGGCTTGAAGCTGACCTGTGTC	1393
Oy	2444	TCACACGCTTCACCAACCACTTCACAGTGCAGTGTCTGGCAAGGCGAGTCAAGAGGA	2503
Db	1392	TCACATGCTCTCACACCACTTCACAGTGCAGTGTCTGGCAAGGCGAGTCAAGAGGA	1333
Oy	2504	CAGGACCAAGGGGCTTTGGGAACAATCCAGACACATCCAAACCCAAAGCGGTGGAACCCACGC	2563
Db	1332	CAGGACCAAGGGGCTTTGGGAACAATCCAGACACATCCAAACCCAAAGCGGTGGAACCCACGC	1273
Oy	2564	CCGGCCCTGGTATCCACAGAGTGGACATGAGCTCCCGCTCTCTGACAGTCCAGCGTCC	2623
Db	1272	CCGGCCCTGGTATCCACAGAGTGGACATGAGCTCCCGCTCTCTGACAGTCCAGCGTCC	1213
Oy	2624	TCACAGAAACAACAAGGCTGTGCTCCGGAACCTTCTTCAGAACCTTCTCCACGTC	2683
Db	1212	TCACAGAAACAACAAGGCTGTGCTCCGGAACCTTCTTCAGAACCTTCTCCACGTC	1153
Oy	2684	CACCTTGGGATGCAGAATGACAGCGAGTAAAGAACCCCTCCACCGGCTTGACCTTGCTGC	2743
Db	1152	CACCTTGGGATGCAGAATGACAGCGAGTAAAGAACCCCTCCACCGGCTTGACCTTGCTGC	1093
Oy	2744	AGTAAAGTTATGATGAGGCGCTGCTCTCCGGGCGCTGGAAGTGGACCAATCAGTTGCTTT	2803
Db	1092	AGTAAAGTTATGATGAGGCGCTGCTCTCCGGGCGCTGGAAGTGGACCAATCAGTTGCTTT	1033
Oy	2804	GCTGACCCCTCGAGAACACCGCCGACAGTGTGGCTGAGACAGCTGGCGGGGGGCC	2863
Db	1032	GCTGACCCCTCGAGAACACCGCCGACAGTGTGGCTGAGACAGCTGGCGGGGGGCC	973
Oy	2864	CCAAGTCGCGCGGCTTCACGCCACCAACAGCTGTGTGAAGTCAAGGCCAACCTCC	2923
Db	972	CCAAGTCGCGCGGCTTCACGCCACCAACAGCTGTGTGAAGTCAAGGCCAACCTCC	913
Oy	2924	AGCACTGATCTGATGAAGGAGCTAAAGAACTCCTCCTCGTGTGTTGAAAGAGGTTGC	2983
Db	912	AGCACTGATCTGATGAAGGAGCTAAAGAACTCCTCCTCGTGTGTTGAAAGAGGTTGC	853
Oy	2984	GATTGTCAATTCGTAAACATTCATCTCCATTTTAAAAAAGGTTTCTGACGGGCC	3043
Db	852	GATTGTCAATTCGTAAACATTCATCTCCATTTTAAAAAAGGTTTCTGACGGGCC	794
Oy	3044	ACGGCCGAGCGCGGTGAGGCGTGTGTCATGATAGCCTGGGCCCCGGGCTTCCGTGCG	3103
Db	793	ACGGCCGAGCGCGGTGAGGCGTGTGTCATGATAGCCTGGGCCCCGGGCTTCCGTGCG	734
Oy	3104	CCTGCGCGGAGGTCCTTGGGGACCACTCTCTGCGTTCAATTTCAGTCGACGTGAC	3163
Db	733	CCTGCGCGGAGGTCCTTGGGGACCACTCTCTGCGTTCAATTTCAGTCGACGTGAC	674
Oy	3164	AGAAAGCACTACCAACAATAAAGCTTTCCTGAAAGCAGA	3202
Db	673	AGAAAGCACTACCAACAATAAAGCTTTCCTGAAAGCAGA	635
RESULT 14			
AK096217/c			
LOCUS AK096217			
DEFINITION Homo sapiens cDNA, FLJ38898 f1s, clone NT2MB2000374.			
ACCESSION AK096217			
VERSION AK096217.1			
KEYWORDS oligo capping; f1s (full insert sequence).			
SOURCE Homo sapiens (human)			







misc_feature	/note="assembly_name:Contig19"	36600. .40210
misc_feature	/note="assembly_name:Contig20"	40311. .43308
misc_feature	/note="assembly_name:Contig21"	43409. .46757
misc_feature	/note="assembly_name:Contig22"	46858. .51135
misc_feature	/note="assembly_name:Contig23"	51236. .55293
misc_feature	/note="assembly_name:Contig24"	55394. .62010
misc_feature	/note="assembly_name:Contig25"	62111. .68999
misc_feature	/note="assembly_name:Contig26"	69100. .74515
misc_feature	/note="assembly_name:Contig27"	74616. .80987
misc_feature	/note="assembly_name:Contig28"	81088. .87983
misc_feature	/note="assembly_name:Contig29"	88084. .98428
misc_feature	/note="assembly_name:Contig30"	98529. .107911
misc_feature	/note="assembly_name:Contig31"	108012. .117076
misc_feature	/note="assembly_name:Contig32"	117177. .128937
misc_feature	/note="assembly_name:Contig33"	129038. .138066
misc_feature	/note="assembly_name:Contig34"	138167. .149334
misc_feature	/note="assembly_name:Contig35"	149435. .165184
misc_feature	/note="assembly_name:Contig37"	165285. .181704
misc_feature	/note="assembly_name:Contig38"	181805. .183334
misc_feature	/note="assembly_name:Contig8"	183435. .185061
misc_feature	/note="assembly_name:Contig9"	

## ORIGIN

Query Match	38.5%	Score 1236.4;	DB 2;	Length 185061;
Best Local Similarity	92.1%;	Pred. No. 0;		
Matches 1273; Conservative	0;	Mismatches 106;	Indels 3;	Gaps 3;

Qy	1821	ATATCAATTATGTGAAGGACATGACAGACACCAGCTTTGAGGGTGTGGGGTAGG	1880
Db	51236	AA	51177
Qy	1881	ACTGAGGACAGCCCACTGGAGAACAGACTGACGCTGGCCATGGCTGTTTTCCAAAGA	1940
Db	51176	NN	51117
Qy	1941	TCAGTTCCTGAGAGGAAGGCTCTGGCCCTGACTCCGCTGTGTCCGAGACACGTGCTG	2000
Db	51116	TCAGTTCCTGAGAGGAGGCTCTGGCCCTGACTCCGCTGTGTCCGAGACACGTGCTG	51057
Qy	2001	ACCGCACCCTGGCCGCTGTAGTTCCTTGGGCTGGAGGTGTCTGTGAGACACCTTG	2060
Db	51056	ACCGCACCCTGGCCGCTGTAGTTCCTTGGGCTGGAGGTGTGTGTGAGACACCTTG	50997
Qy	2061	CCCTCAGACACAGAGGCTGAGGCACTTCTGACGTCCACGCTGAACATGGGAAACAACCTG	2120
Db	50996	CCCTCAGACACAGAGGCTGAGGCACTTCTGACGTCCACGCTGAACATGGGAAACAACCTG	50937
Qy	2121	AAAAGCAGCAGGAGCTCCCGGTGAGGAGGCTGTGTCGTGCTGCTCCATGACCACTT	2180
Db	50936	AAAAGCAGGAGGAGCTCCCGGTGAGGAGGCTGTGTCGTGCTGCTCCATGACCACTT	50878
Qy	2181	CCTCTGCTGAAATATTACTGCTTGAATCTGAGCAGATTCGGGGTTTATAAAATGCTTT	2240

Db	50877	CCCTTCGTGAATAATTACTGCTGTAATCTGAGACGAACTGGGGGTTTAAACCTGCT	50818
Qy	2241	TTTATCTGAGACAAACGGGTTTGGAAATTAAGTCGTCCTTTTCCCACTCCCAAGCTG	2300
Db	50817	TTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCGTCCTTTTCCCACTCCCAAGCTG	50758
Qy	2301	CTCAAGCATTTCCACCGGGCCCCCTGGGCTTGGGGAACAGGGTAGTGTAACTCCGATCCAG	2360
Db	50757	CTCAAGCATTTCCACCGGGCCCCCTGGGCTTGGGGAACAGGGTAGTGTAACTCCGATCCAG	50638
Qy	2361	GGCCTAGCCCTGACACAGAGTGCGTTCCCGTATCCCGTGGGAAAAAGCCCTGCCACAGC	2420
Db	50697	GGCCTAGCCCTGACACAGAGTGCGTTCCCGTATCCCGTGGGAAAAAGCCCTGCCACAGC	50638
Qy	2421	GGGCTTGAAGTGGGCTGTGTCTCTCCACCGCTGCAACCACTCCAGATGGAATGC	2480
Db	50637	GGGCTTGAAGTGGGCTGTGTCTCTCCACCGCTGCAACCACTCCAGATGGAATGC	50578
Qy	2481	TGGGCAAGGGCAGCCTTAAGAGGACAGAGACAGAGCGCTTGGCAAGACATCAGACACCA	2540
Db	50577	TGGGCAAGGGCAGCCTTAAGAGGACAGAGACAGAGCGCTTGGCAAGACATCAGACACCA	50518
Qy	2541	ACCCAAAGCGGTGAGACCCCAAGCGCCGGCCGTGTATCCAGAGGTGCACTGACGCTCC	2600
Db	50517	ACCCAAAGCGGTGAGACCCCAAGCGCCGGCCGTGTATCCAGAGGTGCACTGACGCTCC	50458
Qy	2601	CCGCTCTCTGAGAGTTCACAGGCTCTCTACAGAAACACAGGGCGCTGTGCTCCGAGCCTTC	2660
Db	50457	CCGCTCTCTGAGAGTTCACAGGCTCTCTACAGAAACACAGGGCGCTGTGCTCCGAGCCTTC	50398
Qy	2661	TTGAGACCCCTTCTCTCACAGTGCCCACTTGGAGTGACAAATGACAGCGAGACTAGACCCC	2720
Db	50397	TTGAGACCCCTTCTCTCACAGTGCCCACTTGGAGTGAGAAATGACAGCGAGACTAGACCCC	50338
Qy	2721	TTCACGGGCTTGCACCTCGGCTGACAGTAAAGTTACGTGAGGCTGTCTTCGGGGCTTGA	2780
Db	50337	TTCACGGGCTTGCACCTCGGCTGACAGTAAAGTTACGTGAGGCTGTCTTCGGGGCTTGA	50278
Qy	2781	AGTGAGCAGCATCAGTTGTCTTGTCTGACACCCCTCGAGAGCAACGCGCGACAGAGTGAGC	2840
Db	50277	AGTGAGCAGCATCAGTTGTCTTGTCTGACACCCCTCGAGAGCAACGCGCGACAGAGTGAGC	50218
Qy	2841	TGAGACAGCTGCGCGCGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACACAGCTGTT	2900
Db	50217	TGAGACAGCTGCGCGGGGGGCCCCCAAGCTGCGCGCTCCAGCCCAACACAGCTGTT	50158
Qy	2901	GCTGAAGTCAAGGCAACCTCCCGAGACATGAGATCTGAATACGGCTAAGAACCTCCCTTC	2960
Db	50157	GCTGAAGTCAAGGCTCCCTCCCTCCAGACATGAGATCTGAATACGGCTAAGAACCTCCCTTC	50098
Qy	2961	CTCTGTTTGGAAAGCAGTTGCGGTGTGTCAATTCGTACATTCATCTCCATTTTTTA	3020
Db	50097	CTCTGTTTGGAAAGCAGTTGCGGTGTGTCAATTCGTACATTCATCTCCATTTTTT	50039
Qy	3021	AAAAAGGTTTCTTGACGGGCCCAACGGGCCGAGACCGCGGTGACGTCGTGTGCAATGAC	3080
Db	50038	AAAAAGGTTTCTTGACGGGCCCAACCGCGGCCGAGCGCGGTGACGTCGTGTGCAATGAC	49980
Qy	3081	CTGGGCGCCGGGGCTTCCCGTGCGCCTCTGCGAGAGTGCTTCTTGGGCAACCACTCCCTGC	3140
Db	49979	CTGGGCGCCGGGGCTTCCCGTGCGCCTCTGCGAGAGTGCTTCTTGGGCAACCACTCCCTGC	49920
Qy	3141	GTTTCATTTTGCACTGCATCTGTACAGAAAGCACTCAACAAATAAACCTTTCCTGAAAGCA	3200
Db	49919	GTTTCATTTTGCACTGCATCTGTACAGAAAGCACTCAACAAATAAACCTTTCCTGAAAGCA	49860
Qy	3201	GA 3202	
Db	49859	GA 49858	

Search completed: July 13, 2004, 12:39:19  
Job time : 12312 secs



***This Page Blank (uspto)***



CC respectively. The sequences can be used to treat and prevent disorders  
CC associated with altered expression or activity of HCAP comprising  
CC administering a composition comprising the polypeptide or an antagonist  
CC to a patient (claimed). The human chaperone proteins are also useful for  
CC the diagnosis, treatment or prevention of neurodegenerative, metabolic,  
CC developmental, autoimmune/inflammatory disorders and cell proliferative  
CC disorders including cancer

**SQ** Sequence 3213 BP; 771 A; 876 C; 894 G; 672 T; 0 U; 0 Other;

Query Match	100.0%	Score 3213;	DB 3;	Length 3213;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3213; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	GAAAGTTGGGAAGATGAGCGACGGCCCTTGAAGGAGAGAGAGCTGGACAAATGAACATATT	60
Db	1	GAAAGTTGGGAAGATGAGCGACGGCCCTTGAAGGAGAGAGAGCTGGACAAATGAACATATT	60
Qy	61	ACTGTTGCTGAACGTGCGCAGGGAGGCCCTTCTTGAAGAGCTGAAACCTGCTACCGGA	120
Db	61	ACTGTTGCTGAACGTGCGCAGGGAGGCCCTTCTTGAAGAGCTGAAACCTGCTACCGGA	120
Qy	121	GGCTCTGTATGCTCTACATCCAGACAGACACAGAGCCAGAGCTCAATGCACAGCGG	180
Db	121	GGCTCTGTATGCTCTACATCCAGACAGACACAGAGCCAGAGCTCAATGCACAGCGG	180
Qy	181	AACGACGTGTTAACCTTGTTCACACAGGCTTAAGAAATGATGTAACCCCAAAACAAGG	240
Db	181	AACGACGTGTTAACCTTGTTCACACAGGCTTAATGAATGCTTAATGTAACCCCAAAACAAGG	240
Qy	241	CCATCTATGATATATATGGAAGAGAGACTGGAATGGAAGATGGAGGTTGTGAAA	300
Db	241	CCATCTATGATATATATGGAAGAGAGACTGGAATGGAAGATGGAGGTTGTGAAA	300
Qy	301	GGAGGAGAACCCCTGCTGAAATTCCGAGAGAGATTGAGCGCGTCGAGAGAGAGAGAG	360
Db	301	GGAGGAGAACCCCTGCTGAAATTCCGAGAGAGATTGAGCGCGTCGAGAGAGAGAGAG	360
Qy	361	AGAGAGATTGCAGACGCAACCAATCCCAAGGAGACGATCAGCGTTGAGTAGATCCA	420
Db	361	AGAGAGATTGCAGACGCAACCAATCCCAAGGAGACGATCAGCGTTGAGTAGATCCA	420
Qy	421	CCGACCTTTTGAATCCCTATGATGAGAGAGATGAAGATGTGTCGGGAGTAGCTTCCG	480
Db	421	CCGACCTTTTGAATCCCTATGATGAGAGAGATGAAGATGTGTCGGGAGTAGCTTCCG	480
Qy	481	AGATTGAATTTAATAAATGACATATCCCATGCTCATTGAGGACCCCTTGAACACGACAG	540
Db	481	AGATTGAATTTAATAAATGACATATCCCATGCTCATTGAGGACCCCTTGAACACGACAG	540
Qy	541	ACACAGCCATCTCTCTGGAAGCTCTCAACCCAGATGGAATGAGAGAGGTTCCATT	600
Db	541	ACACAGCCATCTCTCTGGAAGCTCTCAACCCAGATGGAATGAGAGAGGTTCCATT	600
Qy	601	ACCTTGGGCTCAGACGAGTAACTTGGGGCAAGGAGTGGGAGATTGGAATTTGGAACGTG	660
Db	601	ACCTTGGGCTCAGACGAGTAACTTGGGGCAAGGAGTGGGAGATTGGAATTTGGAACGTG	660
Qy	661	GAGACCTACAGGGGCTTTGTTCGATCTCAAGCTGTTCGTAATCTCAACCAAGATGCT	720
Db	661	GAGACCTACAGGGGCTTTGTTCGATCTCAAGCTGTTCGTAATCTCAACCAAGATGCT	720
Qy	721	TTTGGAACAATACTGTGCTTGCAGATTTTCAATCCCGTGGAAATCCGACCCGGCCTGACCA	780
Db	721	TTTGGAACAATACTGTGCTTGCAGATTTTCAATCCCGTGGAAATCCGACCCGGCCTGACCA	780
Qy	781	CTGTCTTAGCTCCGAACTTAGACAGAGAACCCGTGGCTACCTGACGTGGAGTGGGGTA	840
Db	781	CTGTCTTAGCTCCGAACTTAGACAGAGAACCCGTGGCTACCTGACGTGGAGTGGGGTA	840
Qy	841	TCCAGTCAAGCATGAACACTAGCATGTGTCGAGACATTAACCAAGCATTTCACTGTGG	900
Db	841	TCCAGTCAAGCATGAACACTAGCATGTGTCGAGACATTAACCAAGCATTTCACTGTGG	900

QY	901	CCCTGCACTGGGAAATCCCTCACTCTTTGGACATGATACAGACAAATTCCAAG	960
Db	901	CCCTGCACTGGGAAATCCCTCACTCTTTGGACATGATACAGACAAATTCCAAG	960
QY	961	ATGACGATCAACACTCGTGTGAAGGATCCCTCAAGCAGCGTTCTTTGGACGGTGTGG	1020
Db	961	ATGACGATCAACACTCGTGTGAAGGATCCCTCAAGCAGCGTTCTTTGGACGGTGTGG	1020
QY	1021	AGTACGGAGCTGAAGAGAAATCTCCAGGCAACAGCGTTTGGGTGCAGCTGTCAAGCTTG	1080
Db	1021	AGTACGGAGCTGAAGAGAAATCTCCAGGCAACAGCGTTTGGGTGCAGCTGTCAAGCTTG	1080
QY	1081	GAGTTCCACAGGGGGTTTCTCTCAAGTCAAGCTCAACAGGCGCAAGTCAACATCTCT	1140
Db	1081	GAGTTCCACAGGGGGTTTCTCTCAAGTCAAGCTCAACAGGCGCAAGTCAACATCTCT	1140
QY	1141	TCCCTATTCACTTGAACGACCGCTTCTGCCAGCGCATATGTTCTATGCCACCGTGGGGC	1200
Db	1141	TCCCTATTCACTTGAACGACCGCTTCTGCCAGCGCATATGTTCTATGCCACCGTGGGGC	1200
QY	1201	CTCTAAGTGTCTACTTGTGCCATGCAACCGTCTGATCATCAACCATTACTCAGGGCTCAGA	1260
Db	1201	CTCTAAGTGTCTACTTGTGCCATGCAACCGTCTGATCATCAACCATTACTCAGGGCTCAGA	1260
QY	1261	AAGAGAAGGAATTTGAGAGACAGAGAGGAAACGCGCCACCGATGTGCTGCAGAAAGAC	1320
Db	1261	AAGAGAAGGAATTTGAGAGACAGAGAGGAAACGCGCCACCGATGTGCTGCAGAAAGAC	1320
QY	1321	AAGAGCGGAGTCCGCTGTCCGCGTGATGACGAATCTGTCCGAGGATTAATTGAGCAG	1380
Db	1321	AAGAGCGGAGTCCGCTGTCCGCGTGATGACGAATCTGTCCGAGGATTAATTGAGCAG	1380
QY	1381	AAGAGTCCAGATGGGCGCTCATCATGCGCAATGCGTGGTACGGGAAGTTTGTCAATGCA	1440
Db	1381	AAGAGTCCAGATGGGCGCTCATCATGCGCAATGCGTGGTACGGGAAGTTTGTCAATGCA	1440
QY	1441	AGAGCAGAGAGAGCGAGAAAGGTGAAGTGAATGACGTGACGTGCGCCCTGCAGTGCCTGG	1500
Db	1441	AGAGCAGAGAGAGCGAGAAAGGTGAAGTGAATGACGTGACGTGCGCCCTGCAGTGCCTGG	1500
QY	1501	TGAAGGACTCGAACCTCATCTCTCAACGAGCGCTCAAGGCGTGGGCTGCTGGCTTTATG	1560
Db	1501	TGAAGGACTCGAACCTCATCTCTCAACGAGCGCTCAAGGCGTGGGCTGCTGGCTTTATG	1560
QY	1561	ACCGGTGTGTGGGGGAGAGAGAAACCTGAAAGTGTCTATCAATTCGCGGCGCTCTGC	1620
Db	1561	ACCGGTGTGTGGGGGAGAGAGAAACCTGAAAGTGTCTATCAATTCGCGGCGCTCTGC	1620
QY	1621	ATCAGGTGATGTGTCTGGAACGTGAAGGCGCTCGGATACCAAGACATCCCAACGATCG	1680
Db	1621	ATCAGGTGATGTGTCTGGAACGTGAAGGCGCTCGGATACCAAGACATCCCAACGATCG	1680
QY	1681	ATACAGATGATAAATCGCCCAAGAACCAAGATTTTAAAGGCGCAAAAAATCTTTTCT	1740
Db	1681	ATACAGATGATAAATCGCCCAAGAACCAAGATTTTAAAGGCGCAAAAAATCTTTTCT	1740
QY	1741	GGGAGTCTACAAATTTGGAAATGAAAAAACCCAGACATCAAGATTTTATTTATATTA	1800
Db	1741	GGGAGTCTACAAATTTGGAAATGAAAAAACCCAGACATCAAGATTTTATTTATATTA	1800
QY	1801	TTATTTATGAAGGTGTGATCCATTTATCAATTAATGGAAGGACATGACAGACACCCACCTT	1860
Db	1801	TTATTTATGAAGGTGTGATCCATTTATCAATTAATGGAAGGACATGACAGACACCCACCTT	1860
QY	1861	TTGAGGGGTGTGGGGGTAGACTGAGACAGCCCACTGGGAACAGACTGACGCTGGCC	1920
Db	1861	TTGAGGGGTGTGGGGGTAGACTGAGACAGCCCACTGGGAACAGACTGACGCTGGCC	1920
QY	1921	CATGGCTGTTTCCCAAGATCAAGTCTCTGAGGGAAGGGCTGTGGCTCTGACTTCGCTG	1980
Db	1921	CATGGCTGTTTCCCAAGATCAAGTCTCTGAGGGAAGGGCTGTGGCTCTGACTTCGCTG	1980

QY 1981 TGTCCTGAGACACAGTGTCTGACCCGACGCCGCCCTGTAGTCTTCTGAGTGGCTCTGGA 2040  
 DB 1981 TGTCCTGAGACACAGTGTCTGACCCGACGCCGCCCTGTAGTCTTCTGAGTGGCTCTGGA 2040  
 QY 2041 GGTGTCTGTGAGACACCTTGTCTTCAACAAGAGCTGAGCCACTTCTGAGTCTCAAGC 2100  
 DB 2041 GGTGTCTGTGAGACACCTTGTCTTCAACAAGAGCTGAGCCACTTCTGAGTCTCAAGC 2100  
 QY 2101 TGAACATGTGGGAAACAACCTGAAAGAGAGGCTCCCGGTCAAGGAGCCTCTGTCTGTG 2160  
 DB 2101 TGAACATGTGGGAAACAACCTGAAAGAGAGGCTCCCGGTCAAGGAGCCTCTGTCTGTG 2160  
 QY 2161 CTGGCTTCCATGACCAACCTCTCTGTGTAATATTACTCTTGAATCTGGAAGATT 2220  
 DB 2161 CTGGCTTCCATGACCAACCTCTCTGTGTAATATTACTCTTGAATCTGGAAGATT 2220  
 QY 2221 GCGGGTTTATMAAACTGTCTTTTATCTGAGAAACAAGGTTTGGAAATTAGTGTCTTT 2280  
 DB 2221 GCGGGTTTATMAAACTGTCTTTTATCTGAGAAACAAGGTTTGGAAATTAGTGTCTTT 2280  
 QY 2281 TTTTCCCACTCCGAGAGTGTCTCAAGTATCTCAACCGGCCCTCCGCTTGGGACAGGCT 2340  
 DB 2281 TTTTCCCACTCCGAGAGTGTCTCAAGTATCTCAACCGGCCCTCCGCTTGGGACAGGCT 2340  
 QY 2341 AGTGTAACTCCGATCCGAGAGGCTTAGCCCTGACACAAGTGGCTTCCGATCTCCGCTG 2400  
 DB 2341 AGTGTAACTCCGATCCGAGAGGCTTAGCCCTGACACAAGTGGCTTCCGATCTCCGCTG 2400  
 QY 2401 GAAAAAGCGCTGCAACGAGCGGGCTTGAAGCTGGCCCTGTCTTCCATCCGCTGACAC 2460  
 DB 2401 GAAAAAGCGCTGCAACGAGCGGGCTTGAAGCTGGCCCTGTCTTCCATCCGCTGACAC 2460  
 QY 2461 CCACCTCCAGAGTGTCAAGTGTGGGCAAGGGAGCTCAAGAGAGAGACAGAGGCTTGG 2520  
 DB 2461 CCACCTCCAGAGTGTCAAGTGTGGGCAAGGGAGCTCAAGAGAGAGACAGAGGCTTGG 2520  
 QY 2521 CAAGACATCAAGACACACCAACCAAGAGGCTGAGACCCAGAGGCTTGGTATCCCA 2580  
 DB 2521 CAAGACATCAAGACACACCAACCAAGAGGCTGAGACCCAGAGGCTTGGTATCCCA 2580  
 QY 2581 GAGAGTGGCACTGAGACTCCCGCTCTCTGACAGTCTCAAGCTCTTCAACAGAAACAAGG 2640  
 DB 2581 GAGAGTGGCACTGAGACTCCCGCTCTCTGACAGTCTCAAGCTCTTCAACAGAAACAAGG 2640  
 QY 2641 CCTGTGCTCCGAGACCTTCTTCAAGACCTTCTTCAAGTGTCCACTTGGGATCAGAAAT 2700  
 DB 2641 CCTGTGCTCCGAGACCTTCTTCAAGACCTTCTTCAAGTGTCCACTTGGGATCAGAAAT 2700  
 QY 2701 GCAGCGAGAGTAGAGCCCTCTCAACGAGCTGTGAACCTGAGTAAAGTTAGCTGAGG 2760  
 DB 2701 GCAGCGAGAGTAGAGCCCTCTCAACGAGCTGTGAACCTGAGTAAAGTTAGCTGAGG 2760  
 QY 2761 CTTGTCTCTCGGGGCTCTGAAAGTGGACCACTAGTGTCTTCTGCTGACCCCTCGAGCA 2820  
 DB 2761 CTTGTCTCTCGGGGCTCTGAAAGTGGACCACTAGTGTCTTCTGCTGACCCCTCGAGCA 2820  
 QY 2821 AGCCGCGGACAGGTGTGTGTGAGACAGCTGTGGCGGGGGGCCCCAGAGCTGCGCGGCT 2880  
 DB 2821 AGCCGCGGACAGGTGTGTGTGAGACAGCTGTGGCGGGGGGCCCCAGAGCTGCGCGGCT 2880  
 QY 2881 CCAGCCCAACCCAGAGCTGTGTGTAAGTCAAGGCCAACCCTCCAGCACTGTGTATCTGAGT 2940  
 DB 2881 CCAGCCCAACCCAGAGCTGTGTGTAAGTCAAGGCCAACCCTCCAGCACTGTGTATCTGAGT 2940  
 QY 2941 AACGGCTAAGAACTCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3000  
 DB 2941 AACGGCTAAGAACTCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3000  
 QY 3001 ACATTTCATCTTCATTTTAAAGAGTTTCTGACGCGGCCCAACGCGGCCGAGCGGGGT 3060  
 DB 3001 ACATTTCATCTTCATTTTAAAGAGTTTCTGACGCGGCCCAACGCGGCCGAGCGGGGT 3060  
 QY 3061 GAGGCTGTGTGTGATGAGCTGTGGGCCCTCCGCTTCCGCTGTGCGGACAGTGTCT 3120

DB 3061 GAGGCTGTGTGTGATGAGCTGTGGGCCCTCCGCTTCCGCTGTGCGGACAGTGTCT 3120  
 QY 3121 TCTGGGACACCATCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180  
 DB 3121 TCTGGGACACCATCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180  
 QY 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213  
 DB 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213  
 RESULT 2  
 ID ABS64608  
 ID ABS64608 standard; cDNA; 3230 BP.  
 AC ABS64608;  
 XX 15-NOV-2002 (first entry)  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human cDNA probe selectively hybridising to foetal cell mRNA #23.  
 XX  
 KW Human; se; foetal liver myeloid cell; probe; erythroblast;  
 KW foetal abnormality; maternal blood; differential display;  
 KW chromosomal abnormality; single gene disorder; aneuploidy;  
 KW nucleotide triplet expansion disorder; trisomy 13; trisomy 21;  
 KW Klinefelter syndrome; spina bifida; sickle cell anaemia; thalassemia;  
 KW Marfan syndrome; Duchenne muscular dystrophy; cystic fibrosis;  
 KW Fragile X syndrome; Friedreich's ataxia; myotonic dystrophy;  
 KW Huntington's disease; sex determination.  
 KW  
 OS Homo sapiens.  
 XX  
 PN W0200255985-A2.  
 XX  
 PD 18-JUL-2002.  
 XX  
 PF 01-NOV-2001; 2001MO-US045340.  
 XX  
 PR 15-NOV-2000; 2000US-0248882P.  
 XX  
 PA (HOPF ) ROCHE DIAGNOSTICS CORP.  
 XX  
 PI Schueler PA, Xu H, Folz L, Wu X, Sha Y, Nagy A, Mahoney WC;  
 XX  
 DR WPI; 2002-619108/66.  
 DR  
 DR P-PSDB; ABG79328.  
 XX  
 PT Detecting fetal cells in maternal blood sample, useful for diagnosing an  
 PT abnormality in a fetal cell, involves using specific nucleic acid probes  
 PT that hybridize to fetal cell associated RNAs.  
 PT  
 XX  
 PS Claim 109; Page 169-170; 215pp; English.  
 XX  
 CC The invention relates to detecting foetal cells in maternal blood sample,  
 CC involving contacting sample with a first probe or performing expression  
 CC analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to  
 CC that obtained from mature cells to identify RNA/cDNA species useful as  
 CC probe, contacting sample with the probe and identifying if the sample  
 CC comprises a cell that comprises RNA that hybridises to the probe. The  
 CC probes are identified by differential display analysis using mature liver  
 CC cells and foetal liver myeloid cells of less than 22 weeks of gestation.  
 CC Also included are: (1) The probe sequences (appearing as ABS64586-  
 CC ABS64618), an isolated nucleic acid molecule having a sequence which is  
 CC at least 90% identical to the probe or its complement, or having a  
 CC nucleotide sequence identical to at least 15 consecutive nucleotide  
 CC residues of the probe, the encoded proteins from the probe or encoding a  
 CC naturally occurring allelic variant or fragment; (2) a non-mammalian host  
 CC cell containing the probe and (3) an antibody which selectively binds to  
 CC the probe encoded protein. The method is useful for detecting a foetal  
 CC cell (such as erythroblast or trophoblast) in a maternal blood sample,  
 CC which is useful for diagnosing a chromosomal abnormality, single gene  
 CC disorder or nucleotide triplet expansion in the gene, in a foetal cell.



CC For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
CC syndrome), spina bifida, sickle cell anaemia, a thalassemia, Marfan  
CC syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
CC syndrome, Friedreich's ataxia, myotonic dystrophy or Huntington's  
CC disease. The probes are also useful for determination of the sex of a  
CC fetus and for detecting the presence of multiple foetuses at early  
CC stages of pregnancy. The present sequence is a foetal cell specific probe  
CC of the invention  
XX

Sequence 3230 BP; 782 A; 875 C; 897 G; 676 T; 0 U; 0 Other;

Query Match 99.3%; Score 3191.4; DB 6; Length 3230;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 3206; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
Qy 1 GAAAGCTTGGAGAGATGCGCCGCTTGGAGCGAGAGAGAGCTGAGACATGAACTATT 60
Db 7 GAAAGCTTGGAGAGATGCGCCGCTTGGAGCGAGAGAGAGCTGAGACATGAACTATT 66
Qy 61 ACTGCTGTGAAAGTSCGAGAGAGAGCTCTTCTGAAAGAGCTGAAAGCTGCTACCGGA 120
Db 67 ACTGCTGTGAAAGTSCGAGAGAGAGCTCTTCTGAAAGAGCTGAAAGCTGCTACCGGA 126
Qy 121 GAGCTGTGATGCTCTACATCCAGACAGACAGAGCCAGAGCTCAAGTCAAGGCGG 180
Db 127 GAGCTGTGATGCTCTACATCCAGACAGACAGAGCCAGAGCTCAAGTCAAGGCGG 186
Qy 181 AAGAGCTGTTAACTCTGTTTACCAGGCTTATGAAAGTGTAGTGAACCCCAACAGG 240
Db 187 AAGAGCTGTTAACTCTGTTTACCAGGCTTATGAAAGTGTAGTGAACCCCAACAGG 246
Qy 241 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAA 300
Db 247 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAA 306
Qy 301 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGTTTGAAGCGCTGCAGAGAGAGAG 360
Db 307 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGTTTGAAGCGCTGCAGAGAGAGAG 366
Qy 361 AGAGAGATATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 367 AGAGAGATATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Qy 421 CCGACCTTTTGAATCGCTATGATGAGAGATGAAAGTGTCCGAGAGAGCTTTCCG 480
Db 427 CCGACCTTTTGAATCGCTATGATGAGAGATGAAAGTGTCCGAGAGAGCTTTCCG 486
Qy 481 AGATTGAATTTAATAAATSCACATATCCAGTCCATTGAGGCAACCTTTCAGAGCAG 540
Db 487 AGATTGAATTTAATAAATSCACATATCCAGTCCATTGAGGCAACCTTTCAGAGCAG 546
Qy 541 ACAAGCATCTCTCTCGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGGTTTCAATTA 600
Db 547 ACAAGCATCTCTCTCGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGGTTTCAATTA 606
Qy 601 ACTTGGCTCAGACGAGTAATCTTCGCAAGAGAGATGGAGAGTGGAAATTTGAGAGT 660
Db 607 ACTTGGCTCAGACGAGTAATCTTCGCAAGAGAGATGGAGAGTGGAAATTTGAGAGT 666
Qy 661 GAGACCTAAGGGGCTTTGTTGCTCAAGCTGTTCCGTAACTTCACACCAAGATGCT 720
Db 667 GAGACCTAAGGGGCTTTGTTGCTCAAGCTGTTCCGTAACTTCACACCAAGATGCT 726
Qy 721 TTGTGACAACTGCTGCTGCAAGTTTCAATCCGTGAATCCAGCCGGCTGACCA 780
Db 727 TTGTGACAACTGCTGCTGCAAGTTTCAATCCGTGAATCCAGCCGGCTGACCA 786
Qy 781 CTGCTCAGCTCGGAACCTAGACAAAGAACCGTGGGCTACCTGAGTGGCAATGGGGTA 840
Db 787 CTGCTCAGCTCGGAACCTAGACAAAGAACCGTGGGCTACCTGAGTGGCAATGGGGTA 846
Qy 841 TCCAGTACGCTAGAACCTAGACATCGTCCGAGACATTAACACAGCACTTCACTGTG 900
Db 847 TCCAGTACGCTAGAACCTAGACATCGTCCGAGACATTAACACAGCACTTCACTGTG 906
```

```
Db 847 TCCAGTACGCTAGAACCTAGACATCGTCCGAGACATTAACACAGCACTTCACTGTG 906
Qy 901 CCTGAGCTGGAAATCCCTCATCTCTTGGACATGATAGATAGCAAAATTCAG 960
Db 907 CCTGAGCTGGAAATCCCTCATCTCTTGGACATGATAGATAGCAAAATTCAG 966
Qy 961 ATGACATCAGACTCGTGTGAAAGATCCCTCAAGCAGGCTTCTTTGGGACGCTGTG 1020
Db 967 ATGACATCAGACTCGTGTGAAAGATCCCTCAAGCAGGCTTCTTTGGGACGCTGTG 1026
Qy 1021 AGTACGAGCTAGAGAAAGATCTCCAGGACAGGCTTTGGGTGAGCTGCAGCTTG 1080
Db 1027 AGTACGAGCTAGAGAAAGATCTCCAGGACAGGCTTTGGGTGAGCTGCAGCTTG 1086
Qy 1081 GAGTTCACAGGCGCTTCTCTCAAGTCAAGCTCAACAGGCGCATGATCATCTTCT 1140
Db 1087 GAGTTCACAGGCGCTTCTCTCAAGTCAAGCTCAACAGGCGCATGATCATCTTCT 1146
Qy 1141 TCCCTATTCACTTGAAGGACAGCTTCCGAGGCGCATGTTCTATGCAACGTTGGG 1200
Db 1147 TCCCTATTCACTTGAAGGACAGCTTCCGAGGCGCATGTTCTATGCAACGTTGGG 1206
Qy 1201 CTCTAGTGTCTACTTGTGCATGCAACCGTCTGATCATCAACCATACCTCAGGGCTCAG 1260
Db 1207 CTCTAGTGTCTACTTGTGCATGCAACCGTCTGATCATCAACCATACCTCAGGGCTCAG 1266
Qy 1261 AAGAGAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1267 AAGAGAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
Qy 1321 AAGAGCGAGAGTCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGATATTTAGGCGAG 1380
Db 1327 AAGAGCGAGAGTCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGATATTTAGGCGAG 1386
Qy 1381 AAGAGTCCAGAAATGGGCTCTCATGTCATATGCTGTAAGGAAATTTGCAATGACA 1440
Db 1387 AAGAGTCCAGAAATGGGCTCTCATGTCATATGCTGTAAGGAAATTTGCAATGACA 1446
Qy 1441 AAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1447 AAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
Qy 1501 TGAAGAGCTCGAAGCTCATCTCAGAGAGCCTCCAGAGCTGGGCTGCTTTATG 1560
Db 1507 TGAAGAGCTCGAAGCTCATCTCAGAGAGCCTCCAGAGCTGGGCTGCTTTATG 1566
Qy 1561 ACCCGTGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1567 ACCCGTGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
Qy 1621 ATCAGGTGATGGTGTGCAAGTGAAGGCGCTCCGAGATCCAGAGAGTCCAGAGATG 1680
Db 1627 ATCAGGTGATGGTGTGCAAGTGAAGGCGCTCCGAGATCCAGAGAGTCCAGAGATG 1686
Qy 1681 ATACAGATGATTAATCTGCAAGAAACAGATTTTAAAGGCGCGCAAAAATCTTTTCT 1740
Db 1687 ATACAGATGATTAATCTGCAAGAAACAGATTTTAAAGGCGCGCAAAAATCTTTTCT 1746
Qy 1741 GGGAGTCTACAAATTTGGAATGAAAAAACCAGACATCAGATGTTTTTATTAATTA 1800
Db 1747 GGGAGTCTACAAATTTGGAATGAAAAAACCAGACATCAGATGTTTTTATTAATTA 1806
Qy 1801 TTAATTAAGAAGTGTATCAATTAATTAATTAAGAGAGATGACAGACCCAGCTT 1860
Db 1807 TTAATTAAGAAGTGTATCAATTAATTAATTAAGAGAGATGACAGACCCAGCTT 1866
Qy 1861 TTGAGGGTCTGGGGTATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1867 TTGAGGGTCTGGGGTATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
Qy 1921 CATGCTGTTTTCCAGAGATCAGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1927 CATGCTGTTTTCCAGAGATCAGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
```

Oy 1981 TGTCCGAGACACAGTGTGACCCGAGCCGCGCCCTGATGTTCTTGCGTGGGTCTGGA 2040  
 Db 1987 TGTCCGAGACACAGTGTGACCCGAGCCGCGCCCTGATGTTCTTGCGTGGGTCTGGA 2046  
 Oy 2041 GGTGTCTGTGAGACACCTGTGCTTACACAGAGCCGTGAGCACTTCTGAGTCCAGC 2100  
 Db 2047 GGTGTCTGTGAGACACCTGTGCTTACACAGAGCCGTGAGCACTTCTGAGTCCAGC 2106  
 Oy 2101 TGAACATGGGAAACAACCTGAAAGACAGGAGCTCCCGGTGAGGAGCTCTGTCTGTG 2160  
 Db 2107 TGAACATGGGAAACAACCTGAAAGACAGGAGCTCCCGGTGAGGAGCTCTGTCTGTG 2166  
 Oy 2161 CTGGCTTCCATGACACCTCTCTGCTGAAATTTATCTGCTTGAATCTGGAGAGATT 2220  
 Db 2167 CTGGCTTCCATGACACCTCTCTGCTGAAATTTATCTGCTTGAATCTGGAGAGATT 2226  
 Oy 2221 GCGGCTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTTTGAATTAATGTCCTTT 2280  
 Db 2227 GCGGCTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTTTGAATTAATGTCCTTT 2286  
 Oy 2281 TTTCCCTCCTCCAGAGCTGCTCAAGTCAATTCACCGGCTCCCTGTGCTTGAGACAGGT 2340  
 Db 2287 TTTCCCTCCTCCAGAGCTGCTCAAGTCAATTCACCGGCTCCCTGTGCTTGAGACAGGT 2346  
 Oy 2341 AGTGAATCTCCGATCCAGGCGCTGAGCCCTGAGCAAGTGGCTTCCCGTATCCCGGTGG 2400  
 Db 2347 AGTGAATCTCCGATCCAGGCGCTGAGCCCTGAGCAAGTGGCTTCCCGTATCCCGGTGG 2406  
 Oy 2401 GAAAAAGCCCTGACCAACAGCGGCTTGAAGTGGCTGTGTCTTCAACCGCTGACAC 2460  
 Db 2407 GAAAAAGCCCTGACCAACAGCGGCTTGAAGTGGCTGTGTCTTCAACCGCTGACAC 2466  
 Oy 2461 CCACTCTCAAGTGTGAGTCTGTGGCAAGGGCAAGCTCAAGAGACAGACCAAGCGCTTGG 2520  
 Db 2467 CCACCTCCAGAGTGTGAGTCTGTGGCAAGGGCAAGCTCAAGAGACAGACCAAGCGCTTGG 2526  
 Oy 2521 CAAGCATATGACACACCCCAACCCCAAGGCGTGAACCCCAAGCGCGCTGTGTATCCCA 2580  
 Db 2527 CAAGCATATGACACACCCCAACCCCAAGGCGTGAACCCCAAGCGCGCTGTGTATCCCA 2586  
 Oy 2581 GCAGGTGACATGAGAGTCTCCCGCTCTCTGACAGTTCACAGTCTCTTCAACAGAACACAGGG 2640  
 Db 2587 GCAGGTGACATGAGAGTCTCCCGCTCTCTGACAGTTCACAGTCTCTTCAACAGAACACAGGG 2646  
 Oy 2641 CCTGTGCTCGGAGGCTTCTTTCAGACCTTCTTCCAGTGCCTCACTTGGAGTACAGAT 2700  
 Db 2647 CCTGTGCTCGGAGGCTTCTTTCAGACCTTCTTCCAGTGCCTCACTTGGAGTACAGAT 2706  
 Oy 2701 GCAGGGAGGCTAGGACCCCTTCCAGCGGCTGAGACCTCGGCTGCAAGTAAAGTTAGTGG 2760  
 Db 2707 GCAGGGAGGCTAGGACCCCTTCCAGCGGCTGAGACCTCGGCTGCAAGTAAAGTTAGTGG 2766  
 Oy 2761 CCTGTCTCTCGGAGGCTTGAAGTGAAGCCATCAGTTCTCTTGTCTGAGCCCTCGAGCA 2820  
 Db 2767 CCTGTCTCTCGGAGGCTTGAAGTGAAGCCATCAGTTCTCTTGTCTGAGCCCTCGAGCA 2826  
 Oy 2821 AGCGCCGACAGTGTGTGCTGAGACAGCTGCGCGGAGGCGCCCAAGCTGCGCGGCT 2880  
 Db 2827 AGCGCCGACAGTGTGTGCTGAGACAGCTGCGCGGAGGCGCCCAAGCTGCGCGGCT 2886  
 Oy 2881 CCAGCCCAACCAAGCTGTTGCTGAAGTGAAGCCCACTCCCAAGCACTGTATCTGAGT 2940  
 Db 2887 CCAGCCCAACCAAGCTGTTGCTGAAGTGAAGCCCTCCCTCCCAAGCACTGTATCTGAGT 2946  
 Oy 2941 AACGGCTAAGAACTCTTCTCTGTGTTTGAAGAGAGTTCGGGTTGTCCAAATTCGTGA 3000  
 Db 2947 AACGGCTAAGAACTCTTCTCTGTGTTTGAAGAGAGTTCGGGTTGTCCAAATTCGTGA 3006  
 Oy 3001 ACATTCAATCTCAATTTTAAAGGTTTCTGTAGCGGCCCAACGCGCCGAGCGCGGT 3060  
 Db 3007 ACATTCAATCTCAATTTT-AAAAAGGTTTCTGTAGCGGCCCAACGCGCCGAGCGCGGT 3065

Oy 3061 GAGCGTGTGTGATGAGCTGAGGCCCGGAGCTTCCCGTGTGCGCTCTGCGAGAGTGTCT 3120  
 Db 3066 GAGCGTGTGTGATGAGCTGAGGCCCGGAGCTTCCCGTGTGCGCTCTGCGAGAGTGTCT 3125  
 Oy 3121 TCTGGGACCCCATCTCTGCGTTTCATTTGAGTGTGACTGTACAGAAAGGCACTACCA 3180  
 Db 3126 TCTGGGACCCCATCTCTGCGTTTCATTTGAGTGTGACTGTACAGAAAGGCACTACCA 3185  
 Oy 3181 ATAAACCTTTCTGAAAGCGAGAAAAA 3213  
 Db 3186 ATAAACCTTTCTGAAAGCGAGAAAAA 3218  
 RESULT 3  
 AAH14482  
 ID AAH14482 standard; cDNA; 3199 BP.  
 AC AAH14482;  
 XX  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:11986.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 OS  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 11986; 2537pp + Sequence listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification, where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence; where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence and 3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX

Sequence 3199 BP; 759 A; 875 C; 890 G; 675 T; 0 U; 0 Other;

Query Match 98.7%; Score 3171.6; DB 4; Length 3199;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 3188; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
QY 2 AAAGTTGGGAGATGGCGAGCGCTTGAGCGAGGAGCTGAGCAATGAAACTATTA 61
D 3 AAAGTTGGGAGATGGCGAGCGCTTGAGCGAGGAGCTGAGCAATGAAACTATTA 62
QY 62 CTGCTGCTGAGCTGAGCGAGGAGCGCTTCTGAGAGCTGAAAGCTGCTACCGGAG 121
D 63 CTGCTGCTGAGCTGAGCGAGGAGCGCTTCTGAGAGCTGAAAGCTGCTACCGGAG 122
QY 122 GCTCTGATGCTCTACCATCCAGCAAGCAAGAGCCAGAGCTCAAGTCAAGGCGGA 181
D 123 GCTCTGATGCTCTACCATCCAGCAAGCAAGAGCCAGAGCTCAAGTCAAGGCGGA 182
QY 182 ACAGCTGTTTACCTTGTCACAGGCTTATGAAAGTCTTAAAGTCCCAACCAAGGCG 241
D 183 ACAGCTGTTTACCTTGTCACAGGCTTATGAAAGTCTTAAAGTCCCAACCAAGGCG 242
QY 242 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGTGAAAG 301
D 243 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGTGAAAG 302
QY 302 GAGAGAAACCCCTGCTGAAATTTGAGAGAGCTTTGAGCGCTGAGAGAGAGAGAGA 361
D 303 GAGAGAAACCCCTGCTGAAATTTGAGAGAGCTTTGAGCGCTGAGAGAGAGAGAGA 362
QY 362 GAGAGATTCAGAGCGAACAATCCCAAGGGAAGATCAGGTTGAGTAAATGATGCGAC 421
D 363 GAGAGATTCAGAGCGAACAATCCCAAGGGAAGATCAGGTTGAGTAAATGATGCGAC 422
QY 422 CGACCTTTTGTATGCTATGATGAGAGATGTAAGATGTGTCCGCGAGTATGCTTCGCA 481
D 423 CGACCTTTTGTATGCTATGATGAGAGATGTAAGATGTGTCCGCGAGTATGCTTCGCA 482
QY 482 GATTGAAATTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGACAGA 541
D 483 GATTGAAATTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGACAGA 542
QY 542 CACAGCATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGGTTCCATTAA 601
D 543 CACAGCATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGGTTCCATTAA 602
QY 602 CTTTGCGCTCAGACGAGTAACTTCGCGCAAGGGATGGGAGAGTTGGAATTTGAGCTG 661
D 603 CTTTGCGCTCAGACGAGTAACTTCGCGCAAGGGATGGGAGAGTTGGAATTTGAGCTG 662
QY 662 AGACCTTACAGGGGCTTTGTTGGCTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 721
D 663 AGACCTTACAGGGGCTTTGTTGGCTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 722
QY 722 TGAGCAACAACAATGCTGAGTGTTCATCCCGTGAATCGACCCGCGCTGACAC 781
D 723 TGAGCAACAACAATGCTGAGTGTTCATCCCGTGAATCGACCCGCGCTGACAC 782
QY 782 TGTCTTACGCTCGAACCCTAGACAAGAACACGCTGGGCTTACCTGAGTGGAGTAT 841
D 783 TGTCTTACGCTCGAACCCTAGACAAGAACACGCTGGGCTTACCTGAGTGGAGTAT 842
QY 842 CCACTCAGCATGAACTATGATGCTGCTGAGAGACTTAAACCAAGCACTTCACTGTGCG 901
D 843 CCACTCAGCATGAACTATGATGCTGCTGAGAGACTTAAACCAAGCACTTCACTGTGCG 902
QY 902 CCACTCAGCATGAACTATGATGCTGCTGAGAGACTTAAACCAAGCACTTCACTGTGCG 961
D 903 CCACTCAGCATGAACTATGATGCTGCTGAGAGACTTAAACCAAGCACTTCACTGTGCG 962
```

```
QY 962 TGACGATCAGACTCGTGTGAAAATCCCTCAAGAGAGCTTCTTTGGAGCGGTGTGGA 1021
D 963 TGACGATCAGACTCGTGTGAAAATCCCTCAAGAGAGCTTCTTTGGAGCGGTGTGGA 1022
QY 1022 GTACGAGCTGAGAGAAATCTCCAGGCAACGCTTTTGGGTGCACTGTCAAGCTTGG 1081
D 1023 GTACGAGCTGAGAGAAATCTCCAGGCAACGCTTTTGGGTGCACTGTCAAGCTTGG 1082
QY 1082 AGTTCACAGGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCGAGTCAACATCTCTT 1141
D 1083 AGTTCACAGGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCGAGTCAACATCTCTT 1142
QY 1142 CCTTATTCATCTTGAACGACCACTTCTGCGCAGCGCATGTTCTATGCCACCTGGAGGC 1201
D 1143 CCTTATTCATCTTGAACGACCACTTCTGCGCAGCGCATGTTCTATGCCACCTGGAGGC 1202
QY 1202 TCTAGTGTCTACTTTTGGCATGACCCGTGTGATCATCAAAACCATCCTAAGGCTTCAGA 1261
D 1203 TCTAGTGTCTACTTTTGGCATGACCCGTGTGATCATCAAAACCATCCTAAGGCTTCAGA 1262
QY 1262 AGAGAAAGAAATTTGAGAGACAGAGGAAAGCGCCGCAACCGATGCTGACAGAAAGCA 1321
D 1263 AGAGAAAGAAATTTGAGAGACAGAGGAAAGCGCCGCAACCGATGCTGACAGAAAGCA 1322
QY 1322 AGAGCGAGAGTCCGCTGTCGCGCTGAATGCAAGAAATCTGTCCAGAGATTAATTAGGCGAGA 1381
D 1323 AGAGCGAGAGTCCGCTGTCGCGCTGAATGCAAGAAATCTGTCCAGAGATTAATTAGGCGAGA 1382
QY 1382 AGAGTCCAAATATGGGCTCTATCATGCTGCAATGCTGTGTCGCGAGTTTGTCAATGACAA 1441
D 1383 AGAGTCCAAATATGGGCTCTATCATGCTGCAATGCTGTGTCGCGAGTTTGTCAATGACAA 1442
QY 1442 GAGCAGAAAGAGAGAGAGTGAAGTGAATGACGTAATGACGTAATGACGTAATGACGTA 1501
D 1443 GAGCAGAAAGAGAGAGAGTGAAGTGAATGACGTAATGACGTAATGACGTAATGACGTA 1502
QY 1502 GAAAGACTCGAAGCTCATCTCTCAAGAGGCTTCAAGGCTGGGCTGCTGCTTTTATGA 1561
D 1503 GAAAGACTCGAAGCTCATCTCTCAAGAGGCTTCAAGGCTGGGCTGCTGCTTTTATGA 1562
QY 1562 CCCGTGTGTTGGGGGAGAGAAAGAACTTGAATGCTCTATGCTTCCGGGGGCTCTGCA 1621
D 1563 CCCGTGTGTTGGGGGAGAGAAAGAACTTGAATGCTCTATGCTTCCGGGGGCTCTGCA 1622
QY 1622 TCAGGTGATGCTGCTGAGCAGTGAAGGCTTCCGGAATCAAGAGAGTCCCAAGATGCA 1681
D 1623 TCAGGTGATGCTGCTGAGCAGTGAAGGCTTCCGGAATCAAGAGAGTCCCAAGATGCA 1682
QY 1682 TACAGATGATTAACCTGCCAAGAACCAATTTTAAAGCCGCAAAAATCTTTTCTG 1741
D 1683 TACAGATGATTAACCTGCCAAGAACCAATTTTAAAGCCGCAAAAATCTTTTCTG 1742
QY 1742 GGAATCTTAACAATTTGGAAATGAAAACCCAGACATCAGATGTTTATTTATATAT 1801
D 1743 GGAATCTTAACAATTTGGAAATGAAAACCCAGACATCAGATGTTTATTTATATAT 1802
QY 1802 TATTATAGAAAGTGTATACCATTAATATATGAAAGGACATGACAGACACCCAGCTT 1861
D 1803 TATTATAGAAAGTGTATACCATTAATATATGAAAGGACATGACAGACACCCAGCTT 1862
QY 1862 TGAAGGTGCTGGGGGTATGAGACTGAGGAGCCCACTGAGAACCAAGCTGAGCTG 1921
D 1863 TGAAGGTGCTGGGGGTATGAGACTGAGGAGCCCACTGAGAACCAAGCTGAGCTG 1922
QY 1922 ATGAGCTTTTCCCAAGATGAGTTCCTGAGAGAAAGGCTTGGCCCTGATCTGCTGT 1981
D 1923 ATGAGCTTTTCCCAAGATGAGTTCCTGAGAGAAAGGCTTGGCCCTGATCTGCTGT 1982
QY 1982 GTCCGAGACAGAGTCTGACCGGAGCCGCGCTGATGTTCTTTGGCTGGGTCTGAG 2041
D 1983 GTCCGAGACAGAGTCTGACCGGAGCCGCGCTGATGTTCTTTGGCTGGGTCTGAG 2042
QY 2042 GTGTCTGTGAGCAACCTGCTCTCAACAGAGAGGTGAGCACTTCTGAGTCAACGCT 2101
```

Db 2043 GTGTCGTGAGGACCCCTGCCCTCAACAAGAGGTGAGGCACTTTCGACGTCCAACT 2102  
Qy 2102 GAACATGGGAAACAACCTGAAAAAGAGAGAGCCCTCCGGTCAAGGACCTCTGTGTC 2161  
Db 2103 GAACATGGGAAACAACCTGAAAAAGAGAGAGCCCTCCGGTCAAGGACCTCTGTGTC 2162  
Qy 2162 TGGCTTCCCATGACCACTCCCTCGTGAATATTACTGCTGAAATGAGAGGAGATTG 2221  
Db 2163 TGGCTTCCCATGACCACTCCCTCGTGAATATTACTGCTGAAATGAGAGGAGATTG 2222  
Qy 2222 CGGGTTTATAAACTGCTTTTATCTGAAACAAACGGGTTTGAATAATAGTCTGCTTT 2281  
Db 2223 CGGGTTTATAAACTGCTTTTATCTGAAACAAACGGGTTTGAATAATAGTCTGCTTT 2282  
Qy 2282 TTCCCATCTCCAGAGCTGCTCAAGTATTGCACTGGCCCTCGGCTTGGAGACGGGTA 2341  
Db 2283 TTCCCATCTCCAGAGCTGCTCAATTCACCGGCCCTCGGCTTGGAGACGGGTA 2342  
Qy 2342 GTGTAACCTCCGATCCGAGGCGTAAAGCCCTGACACAGGTGGCTTCCGATCCGGTGGG 2401  
Db 2343 GTGTAACCTCCGATCCGAGGCGTAAAGCCCTGACACAGGTGGCTTCCGATCCGGTGGG 2402  
Qy 2402 AAAACGCCCTGCAACCAAGCGGCTTGAAGCTGGCTGTGCTCCCTCAACCGCTGACACAC 2461  
Db 2403 AAAACGCCCTGCAACCAAGCGGCTTGAAGCTGGCTGTGCTCCCTCAACCGCTGACACAC 2462  
Qy 2462 CACCTCCAGAGTGAAGTCTGGGCAAGGCGAGCTCAAGAGAGACAGAGACAGGCGCTTGGC 2521  
Db 2463 CACCTCCAGAGTGAAGTCTGGGCAAGGCGAGCTCAAGAGAGACAGAGACAGGCGCTTGGC 2522  
Qy 2522 AAGCATGAGACACACCAAGGCGTGAAGCCCAAGGCGCGGCGGTGATCCAG 2581  
Db 2523 AAGCATGAGACACACCAAGGCGTGAAGCCCAAGGCGCGGCGGTGATCCAG 2582  
Qy 2582 CAGGTGAGTGAAGTCCCGCTCTGCAAGTCCAGGCTCCTCAACAGAAACACAGAGGC 2641  
Db 2583 CAGGTGAGTGAAGTCCCGCTCTGCAAGTCCAGGCTCCTCAACAGAAACACAGAGGC 2642  
Qy 2642 CTGTGCTCGGAGCCTTCTTCAAGCCTTCTTCTCAAGTCCCACTTGGAGTGAAGATG 2701  
Db 2643 CTGTGCTCGGAGCCTTCTTCAAGCCTTCTTCTCAAGTCCCACTTGGAGTGAAGATG 2702  
Qy 2702 CAGGAGAGTGAAGCCCTCCTCAAGGCTTGAAGCTCCGCTCAAGTAAAGTTACGTAAGGC 2761  
Db 2703 CAGGAGAGTGAAGCCCTCCTCAAGGCTTGAAGCTCCGCTCAAGTAAAGTTACGTAAGGC 2762  
Qy 2762 CTGTCTCTCGGAGCCTTCTTCAAGCCTTCTTCTCAAGTCCCACTTGGAGTGAAGATG 2821  
Db 2763 CTGTCTCTCGGAGCCTTCTTCAAGCCTTCTTCTCAAGTCCCACTTGGAGTGAAGATG 2822  
Qy 2822 GCGCGGACAGGTGCTGAGACAGTGGCGCGGAGGAGCCCAAGCTGCGCGGCTC 2881  
Db 2823 GCGCGGACAGGTGCTGAGACAGTGGCGGTGGGAGGCCCAAGCTGCGCGGCTC 2882  
Qy 2882 CAGCCACCCACAGTGTGTGTAAGTCAAGCTCCCTCCACAGACCTGATCTGAATA 2941  
Db 2883 CAGCCACCCACAGTGTGTGTAAGTCAAGCTCCCTCCACAGACCTGATCTGAATA 2942  
Qy 2942 ACGGCTAAGACCTCTCTCTGATTTGAAAAGCAGTTCGGGTTGCAAACTCTGTA 3001  
Db 2943 ACGGCTAAGACCTCTCTCTGATTTGAAAAGCAGTTCGGGTTGCAAACTCTGTA 3002  
Qy 3002 CATTATCTCAATTTTAAAAAGGTTCTCTGAAGGCGCCACAGGCGCGCGGTG 3061  
Db 3003 CATTATCTCAATTTTAAAAAGGTTCTCTGAAGGCGCCACAGGCGCGCGGTG 3062  
Qy 3062 AGCGTCGTTGACATGAGCTGGGCGCGGCTTCCCTGTGCGCTGAGCCAGAGTGT 3121  
Db 3062 AGCGTCGTTGACATGAGCTGGGCGCGGCTTCCCTGTGCGCTGAGCCAGAGTGT 3121  
Qy 3122 CTGGGACACCATCTCTGCGTTTCAATTTGACAGTGAAGTGAAGGACCTCAACACAA 3181

Db 3122 CTGGGACACCATCTCTGCGTTTCAATTTGACAGTGAAGTGAAGGACCTCAACACAA 3181  
Qy 3182 TAAACCTTTCTGAAAC 3199  
Db 3182 TAAACCTTTCTGAAAC 3199  
RESULT 4  
AACT7352  
ID AACT7352 standard; cDNA; 2991 BP.  
XX  
AC AACT7352;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2907 polynucleotide sequence SEQ ID NO:5813.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; se.  
XX  
OS Homo sapiens.  
XX  
PN MO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000MO-US008621.  
XX  
PR 01-MAR-1999; 99US-0127607P.  
PR 02-APR-1999; 99US-0127636P.  
PR 05-APR-1999; 99US-0127728P.  
PR 30-MAR-2000; 2000US-00540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX P-PSDB; AAB43143.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.  
XX  
PS Claim 5; Page 4978-4980; 5507P; English.  
XX  
CC AACT7446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC antidiabetic; antirheumatic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage degeneration, nocturnal haemoglobinuria, anti-inflammatory diseases; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 2991 BP; 700 A; 817 C; 818 G; 655 T; 0 U; 1 Other;

Query Match 76.4%; Score 2456; DB 3; Length 2991;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 2481; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```
QY 716 ATGCTTTGTGACAACTGCTGCTGCAAGTTTCACTCCGTGGAATCCGACCCGCT 775
DB 505 AAGCTTTGTGACAACTGCTGCTGCAAGTTTCACTCCGTGGAATCCGACCCGCT 564
QY 776 GACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835
DB 565 GACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 624
QY 836 GGGTATCCAGTCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 895
DB 625 GGGTATCCAGTCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 684
QY 896 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955
DB 685 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
QY 956 CCAAGATGAGATCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014
DB 745 CCAAGATGAGATCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 804
QY 1015 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1074
DB 805 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
QY 1075 GCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134
DB 865 GCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924
QY 1135 ACTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1194
DB 925 ACTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 984
QY 1195 TGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1254
DB 985 TGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1044
QY 1255 CTGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1314
DB 1045 CTGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1104
QY 1315 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1374
DB 1105 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1164
QY 1375 AGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1434
DB 1165 AGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1224
QY 1435 ATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
DB 1225 ATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284
QY 1495 GCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
DB 1285 GCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
QY 1555 TTTATGACCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1614
DB 1345 TTTATGACCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1404
QY 1615 TCCTCATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1674
```

```
DB 1405 TCCTCATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1464
QY 1675 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1734
DB 1465 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1524
QY 1735 TTTCTGGAGAGTCTACAAATTTGAAATGAAAAAACCAGATCAGATGTTTTATTTT 1794
DB 1525 TTTCTGGAGAGTCTACAAATTTGAAATGAAAAAACCAGATCAGATGTTTTATTTT 1584
QY 1795 ATTTATTTATTTATGAGAGTGTGATCATTATATATGTTGAAAGGACATGACACCC 1854
DB 1585 ATTTATTTATTTATGAGAGTGTGATCATTATATATGTTGAAAGGACATGACACCC 1644
QY 1855 CAGCTTTGAGAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
DB 1645 CAGCTTTGAGAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1704
QY 1915 CTGGCCATGAGCTGTTTTCCCAAGATCAGTTCTGGAAGAGAGAGAGAGAGAGAGAG 1974
DB 1705 CTGGCCATGAGCTGTTTTCCCAAGATCAGTTCTGGAAGAGAGAGAGAGAGAGAGAG 1764
QY 1975 CCGCTGTGTCCGAGAGACACGTGCTGACCGCAGCCCGCCCTGTGATTTGCTGGG 2034
DB 1765 CCGCTGTGTCCGAGAGACACGTGCTGACCGCAGCCCGCCCTGTGATTTGCTGGG 1824
QY 2035 TCTGAGAGTGTCTGTGAGAGACACCGTCCCTCACAGAGAGAGAGAGAGAGAGAG 2094
DB 1825 TCTGAGAGTGTCTGTGAGAGACACCGTCCCTCACAGAGAGAGAGAGAGAGAGAG 1884
QY 2095 CCAAGCTGAACATGAGAGAAACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
DB 1885 CCAAGCTGAACATGAGAGAAACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1944
QY 2155 GCTGTGTGTGCTTTCCCATGACACCTCTCTCTGCTGAAATATTTAATGCTTGA 2214
DB 1945 GCTGTGTGTGCTTTCCCATGACACCTCTCTCTGCTGAAATATTTAATGCTTGA 2004
QY 2215 CAGATTTGAGAGTTTAAATGCTTTTATCTGAGAAACAAAGAGAGAGAGAGAGAG 2274
DB 2005 CAGATTTGAGAGTTTAAATGCTTTTATCTGAGAAACAAAGAGAGAGAGAGAGAG 2064
QY 2275 GTCTTTTTCCTCCACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
DB 2065 GTCTTTTTCCTCCACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124
QY 2335 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2394
DB 2125 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2184
QY 2395 CCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
DB 2185 CCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2244
QY 2455 CACCAACCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2514
DB 2245 CACCAACCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 2515 GCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2574
DB 2305 GCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2364
QY 2575 TACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2634
DB 2365 TACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2424
QY 2635 CAGAGAGCTGTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2694
DB 2425 CAGAGAGCTGTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2484
QY 2695 CAGATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2754
```

Db	2485	CAGAAATGACGGAGACTAGAACCCCTCCACGGGCTTGAAACCTCGGCTGCACTAAAGATTAC	2544
QY	2755	GTGAGGCTGTCTCTCTGGGGCTTGAAATGGACGACATCAATTTGCTCTTGTCTGACCCCTC	2814
Db	2545	GTGAAGGCTGTCTCTCGGGGCTTGAAATGGACGACATCAATTTGCTCTTGTCTGACCCCTC	2604
QY	2815	GGAGCAACGCGCGCACAGTGGTGGCTAGACAGCTGGCGGGGGGGGCCCAAGTGTGC	2874
Db	2605	GGAGCAACGCGCGCACAGTGGTGGCTAGACAGCTGGCGGGGGGGGCCCAAGTGTGC	2664
QY	2875	CGGCTCCAGCCCAACCAAGCTGTGTGAAGTCAAGCCAACTTCCGACACATGGTAT	2934
Db	2665	CGGCTCCAGCCCAACCAAGCTGTGTGTGAAGTCAAGCCCTCCCAACACATGGTAT	2724
QY	2935	CTGAGTAACGGCTAGAAACCTCTTCCTCTGGTTTTGAAAGCAAGTTGGGTTGTCCAT	2994
Db	2725	CTGAGTAACGGCTAGAAACCTCTTCCTCTGGTTTTGAAAGCAAGTTGGGTTGTCCAT	2784
QY	2995	TCTGTAACTTCAATCTCCATTTTTTAAAAAGTTTCTCTGACGGGCCCAAGCCCGAGC	3054
Db	2785	TCTGTAACTTCAATCTCCATTTTTT -AAAAAGTTTCTCTGACGGGCCCAAGCCCGAGC	2843
QY	3055	CGCGGTGACGCTCGGTGGCATGAGCCCTGGGCCCCCGGCTTCCGCTGGGCTCTTGCAGCA	3114
Db	2844	CGCGGTGACGCTCGGTGGCATGAGCCCTGGGCCCCCGGCTTCCGCTGGGCTCTTGCAGCA	2903
QY	3115	GGTGTCTTGGGACCCCATCTCTGCGTTTCATTTGCACTGACCTGTACAGAAGGCACTC	3174
Db	2904	GGTGTCTTGGGACCCCATCTCTGCGTTTCATTTGCACTGACCTGTACAGAAGGCACTC	2963
QY	3175	ACCAACAATAACCTTTCTGAAAGAGCA	3202
Db	2964	ACCAACAATAACCTTTCTGAAAGAGCA	2991

XX	ADBB63642	standard; cDNA; 2454 BP.
XX	ADBB63642	
XX	ADBB63642	
XX	04-DEC-2003	(first entry)
XX		
DE	Human cDNA encoding clone THYMU20089900.	
XX		
KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;	
KW	tissue regeneration; cell regeneration; membrane protein;	
KW	signal transduction-related protein; transcription-related protein;	
KW	osteoporosis; neurological disease; cancer; tumour.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1008..1421
FT		/*tag= a
FT		/product= "Clone THYMU20089900 protein"
XX		
PN	EP1308459-A2.	
XX		
PD	07-MAY-2003.	
XX		
PE	28-MAR-2002; 2002EP-00007401.	
XX		
PR	05-NOV-2001; 2001JP-00379298.	
PR	25-JAN-2002; 2002US-00350978.	
XX		
PA	(HELI-) HELIX RES INST.	
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.	
XX		
P1	Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,	
P1	Yanemoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,	
XX	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maseho Y;	
XX		

DR WP1: 2003-450961/43.  
 DR P-PSDB: ADB65612.  
 XX  
 XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 PR marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1, Page: 222pp; English.  
 XX  
 XX The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC is based on sequence information supplied by the European Patent Office.  
 XX  
 XX Sequence 2454 BP; 520 A; 707 C; 687 G; 540 T; 0 U; 0 Other;

Query Match	57.6%	Score 1851.6	DB 9	Length 2454
Best Local Similarity	99.7%	Pred. No. 0		
Matches 1865	Conservative 0	Mismatch 4	Indels 1	Gaps 1
QY 1338	GTCCGGCTGATGCAAGATCTGTCCTCGAAGTAATTAGGCGAGAAGATCCAGAAATGGG	1397		
Db 585	GTCCGGCTGATGCAAGATCTGTCCTCGAAGTAATTAGGCGAGAAGATCCAGAAATGGG	644		
QY 1398	CTCATCATCTGCAATGCTTGATCGGGAAGTTTGTCAATGACAAAGACAGAAAGCGAG	1457		
Db 645	CTCATCATCTGCAATGCTTGATCGGGAAGTTTGTCAATGACAAAGACAGAAAGCGAG	704		
QY 1458	AAGGTGAAGGTGATTTGAACGTGACGTGACCTCGAGAGCGCTGTGGAAAGGATCGAACCCT	1517		
Db 705	AAGGTGAAGGTGATTTGAACGTGACGTGACCTCGAGAGCGCTGTGGAAAGGATCGAACCCT	764		
QY 1518	ATCTCTACAGAGGCTCTCAAGGCTGGGCTGCTGGCTTTTATGAACCCGTGTGGGGGAA	1577		
Db 765	ATCTCTACAGAGGCTCTCAAGGCTGGGCTGCTGGCTTTTATGAACCCGTGTGGGGGAA	824		
QY 1578	GAGAAAGAACTGAAAGTGTCTTATCAATTCCGGGGGCTCTGCAATCAGGTGATGTGTCTG	1637		
Db 825	GAGAAAGAACTGAAAGTGTCTTATCAATTCCGGGGGCTCTGCAATCAGGTGATGTGTCTG	884		
QY 1638	GACAGTAGGCGCTCCCGATATCCAAACAGTCCACAGATTCGATACAGATGGATTAAC	1697		
Db 885	GACAGTAGGCGCTCCCGATATCCAAACAGTCCACAGATTCGATACAGATGGATTAAC	944		
QY 1698	GCCAAAGAACAGATTTTAAAGGCGCAAAAAATCTTTTCTGGGAGTCAAAATTTG	1757		
Db 945	GCCAAAGAACAGATTTTAAAGGCGCAAAAAATCTTTTCTGGGAGTCAAAATTTG	1004		
QY 1758	GAATGAAAAAACCCAGACATCGATGTGTTTTATTATATTTATTTATTAAGAGGTGT	1817		
Db 1005	GAATGAAAAAACCCAGACATCGATGTGTTTTATTATATTTATTTATTAAGAGGTGT	1064		
QY 1818	ACCATTTACATTTATGGAAGGACATCGACAGACCCACGCTTTTGAGGCTGTGGGGGT	1877		









QY 1849 ACAACCCAGCTTTTGAGGGGTGCTGGGGGTAGAGCTGAGGAGCCCACTGGGAACGAGC 1908  
| | | | |  
Db 18348 ACAACCCAGCTTTTGAGGGGTGCTGGGGGTAGAGCTGAGGAGCCCACTGGGAACGAGC 18407  
| | | | |  
QY 1909 TGAAGCTGGCCCAATGAGCTGTTTCCCAAGATCAGTTCTGGAAGGAGGCTCTGGCC 1968  
| | | | |  
Db 18408 TGAAGCTGGCCCAATGAGCTGTTTCCCAAGATCAGTTCTGGAAGGAGGCTCTGGCC 18467  
| | | | |  
QY 1969 CTGACTCCGCTGATGCTCCGAGCAAGGTGCTGACCCGAGCCGCGCTGTAAGTTCTTG 2028  
| | | | |  
Db 18468 CTGACTCCGCTGATGCTCCGAGCAAGGTGCTGACCCGAGCCGCGCTGTAAGTTCTTG 18527  
| | | | |  
QY 2029 GCTGGGTCTGAGAGTGTCTGTGAGAGCACTCTGCTTCAACCAAGAGCGTGAACCTTC 2088  
| | | | |  
Db 18528 GCTGGGTCTGAGAGTGTCTGTGAGAGCACTCTGCTTCAACCAAGAGCGTGAACCTTC 18587  
| | | | |  
QY 2089 TGAAGTTCAGAGCTGGAACATGGGAAACAACCTGAAAGCAGGCGCTCCCGTCAAGGA 2148  
| | | | |  
Db 18588 TGAAGTTCAGAGCTGGAACATGGGAAACAACCTGAAAGCAGGCGCTCCCGTCAAGGA 18647  
| | | | |  
QY 2149 GCGCTGCTGCTGCTGCTCCCAAGACCACTCTCTGCTGAATATTTCTGCTTGAAT 2208  
| | | | |  
Db 18648 GCGCTGCTGCTGCTGCTCCCAAGACCACTCTCTGCTGAATATTTCTGCTTGAAT 18707  
| | | | |  
QY 2209 CTGAGAGCAAGTTGCGGGTTTATAAAACTGCTTTTATCTGAGAAACAACGGGTTTGGAAA 2268  
| | | | |  
Db 18708 CTGAGAGCAAGTTGCGGGTTTATAAAACTGCTTTTATCTGAGAAACAACGGGTTTGGAAA 18767  
| | | | |  
QY 2269 TTATGCTCTTTTTCCTCCCACTCCCAAGAGTGTCTCAAGTATTCACCCGCGCTTCGGC 2328  
| | | | |  
Db 18768 TTATGCTCTTTTTCCTCCCACTCCCAAGAGTGTCTCAAGTATTCACCCGCGCTTCGGC 18827  
| | | | |  
QY 2329 TTGGGACAGGGTATGTAATCTCCGATCCAGGAGCTGAGCCCTGACACAGAGTGGCTTCCC 2388  
| | | | |  
Db 18828 TTGGGACAGGGTATGTAATCTCCGATCCAGGAGCTGAGCCCTGACACAGAGTGGCTTCCC 18887  
| | | | |  
QY 2389 GTATCCCGGTGGGAAAAAGCGCTTCCCAACAGCGGGCTTGAAGTGGCTGTGTCTCCAC 2448  
| | | | |  
Db 18888 GTATCCCGGTGGGAAAAAGCGCTTCCCAACAGCGGGCTTGAAGTGGCTGTGTCTCCAC 18947  
| | | | |  
QY 2449 CGGCTGACCACTCCCTCCAGAGTGCATGCTGGGGCAAGGCAAGTCAAGAGACAGGA 2508  
| | | | |  
Db 18948 CGGCTGACCACTCCCTCCAGAGTGCATGCTGGGGCAAGGCAAGTCAAGAGACAGGA 19007  
| | | | |  
QY 2509 CCAGCGCTTGGCAAGCATCAGACACACCAACCCAAAGGCTGAGACCCAGCGCGG 2568  
| | | | |  
Db 19008 CCAGCGCTTGGCAAGCATCAGACACACCAACCCAAAGGCTGAGACCCAGCGCGG 19067  
| | | | |  
QY 2569 CCGTGGTACCCAGAGGTGCACTGAGCTCCCGCTCTGCAAGTTCAGCGTCTTCA 2628  
| | | | |  
Db 19068 CCGTGGTACCCAGAGGTGCACTGAGCTCCCGCTCTGCAAGTTCAGCGTCTTCA 19127  
| | | | |  
QY 2629 GGAACACAGGGCTGTGCTCCGAGAGCTTCCCTTCAAGCCCTTCCCAAGTCCCACTT 2688  
| | | | |  
Db 19128 GGAACACAGGGCTGTGCTCCGAGAGCTTCCCTTCAAGCCCTTCCCAAGTCCCACTT 19187  
| | | | |  
QY 2689 GGGATGCAAGATCAGAGGAGTAGAGCCCTTCAACGAGCTGAGACCTCGGCTGAGTAA 2748  
| | | | |  
Db 19188 GGGATGCAAGATCAGAGGAGTAGAGCCCTTCAACGAGCTGAGACCTCGGCTGAGTAA 19247  
| | | | |  
QY 2749 AGTTAAGTAGAGCTGTCTCTCGAGGCTTGAAGATGGGACCAATCAGTTGCTTGTGA 2808  
| | | | |  
Db 19248 AGTTAAGTAGAGCTGTCTCTCGAGGCTTGAAGATGGGACCAATCAGTTGCTTGTGA 19307  
| | | | |  
QY 2809 CCCCTGGAGCAAGCCCGGACAGGTGCTGAGACAGCTGAGCCCGGGGGGCCCCAAG 2868  
| | | | |  
Db 19308 CCCCTGGAGCAAGCCCGGACAGGTGCTGAGACAGCTGAGCCCGGGGGGCCCCAAG 19367  
| | | | |  
QY 2869 CTGCGCGGCTCTCAGGCCACCAACAGCTTGTGTAAGTCAAGGCCAACCTTCCAGACAC 2928  
| | | | |  
Db 19368 CTGCGCGGCTCTCAGGCCACCAACAGCTTGTGTAAGTCAAGGCCAACCTTCCAGACAC 19427  
| | | | |  
QY 2929 TGGTATCTGAGTAAAGGCTAAAGAACTCTTCTCTGTTTGGAAAAAGCAGTTTGGGTTG 2988  
| | | | |

Db 19428 TGGTATCTGAGTAAAGGCTAAAGAACTCTTCTCTGTTTGGAAAAAGCAGTTGCGGTTG 19487  
| | | | |  
QY 2989 TCCATATCTGTAAATCATTCCTCAATTTTAAAAAAGTTTCTGAGAGGCCCAAGCC 3048  
| | | | |  
Db 19488 TCCATATCTGTAAATCATTCCTCAATTTTAAAAAAGTTTCTGAGAGGCCCAAGCC 19546  
| | | | |  
QY 3049 CCGAGCCGCGGTGAGAGGTGTGTTGATGAGCTGAGCCCGGGGCTTCCCGTGGCGCTCT 3108  
| | | | |  
Db 19547 CCGAGCCGCGGTGAGAGGTGTGTTGATGAGCTGAGCCCGGGGCTTCCCGTGGCGCTCT 19606  
| | | | |  
QY 3109 GCCGAGGTGCTTCTGGGACCCATCTCTGCGTTTCATTTGCACTGAGAGAG 3168  
| | | | |  
Db 19607 GCCGAGGTGCTTCTGGGACCCATCTCTGCGTTTCATTTGCACTGAGAGAG 19666  
| | | | |  
QY 3169 GCATCTACCAATTAACCTTCTCGAAGCAGA 3202  
| | | | |  
Db 19667 GCATCTACCAATTAACCTTCTCGAAGCAGA 19700  
| | | | |  
RESULT 7  
AB054721  
ID AB054721 standard; cDNA; 1454 BP.  
XX  
AC AB054721;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HPAM601 cDNA, SEQ ID NO:601.  
XX  
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
XX ovarian cancer; breast cancer; tumour; reproductive system disorder;  
XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
XX PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
XX inflammatory condition; immune disorder; blood disorder;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disorder; urinary system disorder; drug screening;  
XX gene therapy; chromosome mapping; forensic analysis;  
XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
XX antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200200677-A1.  
XX  
PD 03-JUN-2002.  
XX  
XX 07-JUN-2001; 2001WO-US018569.  
XX  
XX 07-JUN-2000; 2000US-0209467P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PA  
XX Birse CE, Rosen CA;  
XX PT  
XX DR MPI; 2002-147878/19.  
XX  
XX DR P-PSDB; ABP41644.  
XX  
XX  
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX  
PS Claim 1; SEQ ID NO 601; 2922bp; English.  
XX  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
XX ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also  
XX encompasses polypeptides 90% identical and polynucleotides 95% identical  
XX to the sequences of the invention. The invention additionally relates to  
XX recombinant vectors and host cells comprising human ovarian antigen  
XX polynucleotides, antibodies against human ovarian antigens, and the use  
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,  
XX treating, prognosing or preventing various ovary and/or breast-related



PD		21-SEP-2000.
PP		
PF	08-MAR-2000;	2000WO-US005988.
PR	12-MAR-1999;	99US-0124270P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	(ROSE/) ROSEN C A.	
PA	Rosen CA,	Ruben SM;
PI		
DR	WPI; 2000-587513/55.	
DR	P-PsDB; AAB56849.	
XX		
PT	Prostate cancer associated gene sequences, referred to as prostate cancer	
PT	antigens, useful for treatment, prevention, and diagnosis of disorders	
XX	such as prostate cancer.	
PS	Claim 1; Page 957-958; 2338bp; English.	
XX		
CC	AAP1556 to AAP1505 encode the human prostate cancer associated	
CC	proteins, called prostate cancer antigens, given in AAB56363 to AAB57032.	
CC	The prostate cancer antigens can have neuroprotective, cytostatic,	
CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,	
CC	nephrotropic, antileptic, gynaecological and antibacterial activities,	
CC	and can be used in gene therapy. The prostate cancer antigen	
CC	polynucleotides may be used for detection of prostate cancer, chromosome	
CC	identification, as chromosome markers, and for numerous other diagnostic	
CC	or research purposes. The prostate cancer antigens may be used to treat	
CC	disorders such as neural, immune, muscular, reproductive,	
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative	
CC	disorders, wounds, and infectious diseases. AAP1506 to AAP1534 to	
CC	AAB57303 represent sequences used in the exemplification of the present	
CC	invention	
XX		
SQ	Sequence 1451 BP; 308 A; 449 C; 386 G; 304 T; 0 U; 4 Other;	
	Query Match	44.0%; Score 1413.6; DB 3; Length 1451;
	Best Local Similarity	99.5%; Pred. No. 0;
	Matches 1424; Conservative	3; Mismatches 3; Indels 1; Gaps 1;
OY	1783	TGTTTTATTATTTAATTATTAATAGAAAGTGTCACCATTAATATGTGAAGGAC 1842
Db	1	TGTTTTATTATTTAATTATTAATAGAAAGTGGATACCAATTATGTGAAGGAC 60
OY	1843	ATGCAGAACCACCCAGCTTTTAGAGGGTGTCTGGGGGTGAAGGACTGAGGACGCCCATCTGGAA 1902
Db	61	ATGCAGAACCACCCAGCTTTTAGAGGGTGTCTGGGGGTGAAGGACGCCCATCTGGAA 120
OY	1903	CCAGACTCAGCTCGGCCCATGTGCTTTCCCAAAGATCAGTTCTTGAAGGAAAGGCT 1962
Db	121	CCAGACTCAGCTCGGCCCATGTGCTTTCCCAAAGATCAGTTCTTGAAGGAAAGGCT 180
OY	1963	CTGGCCCTGACTCCGCTGTGTCCGAGACAACGTGCTGACCGCAGCCCGCCTGTAG 2022
Db	181	CTGGCCCTGACTCCGCTGTGTCCGAGACAACGTGCTGACCGCAGCCCGCCTGTAG 240
OY	2023	TTCTTTGGCGGGTCTGGAAGGTGTCTGTGAAGAACCTTCCTCCACACAGAGGCTGAGC 2082
Db	241	TTCTTTGGCGGGTCTGGAAGGTGTCTGTGAAGAACCTTCCTCCACACAGAGGCTGAGC 300
OY	2083	CACATTCTGACGTCAACGCTGAACATGGAAAAACAACCTTGAAGGAGGAGGCT 2142
Db	301	CACATTCTGACGTCAACGCTGAACATGGAAAAACAACCTTGAAGGAGGAGGCT 360
OY	2143	CAGGAGGCTCTGCTGTGCTGTGCTTCCATGACCACTTCCTGCTGAATAATTAATCTGC 2202
Db	361	CAGGAGGCTCTGCTGTGCTGTGCTTCCATGACCACTTCCTGCTGAATAATTAATCTGC 420
OY	2203	TTGAATCTGAGAGCAGATTGCGGGTTTATAAAAATGCTTTTATCTGAAGAACAAACGGGTT 2262
Db	421	TTGAATCTGAGAGCAGATTGCGGGTTTATAAAAATGCTTTTATCTGAAGAACAAACGGGTT 480

QY	2265	TGGAATTAATGACGCTTTTTC	CCACATCCGAGAGCGCTCAAGTACTTCCACCGGCCCC	2232
Db	481	TGGAATTAATGACGCTTTTTC	CCACATCCGAGAGCGCTCAAGTACTTCCACCGGCCCC	540
QY	2323	CTCGGCTTGGGAACAGGAGTGTAACTCCGATCCAGGGCTTAGCCCTGACACAGGTGG	2382	
Db	541	CTCGGCTTGGGAACAGGAGTGTAACTCCGATCCAGGGCTTAGCCCTGACACAGGTGG	600	
QY	2383	CTTCCCGTATCCGGTGGGAAAAAGCCCTGCACACAGCGGACTTGAAGCTGTGTCC	2442	
Db	601	CTTCCCGTATCCGGTGGGAAAAAGCCCTGCACACAGCGGACTTGAAGCTGTGTCC	660	
QY	2443	CTCCACCGGCTTGACACACCACTCCGAGATGACAGTGTGGGCAAGGGCAGCTCAAGAG	2502	
Db	661	CTCCACCGGCTTGACACCACTCCGAGATGACAGTGTGGGCAAGGGCAGCTCAAGAG	720	
QY	2503	ACAGAGCAAGGCGCTTGGCAAGACATAGACACACCCAAACCCAAAGGCGTGGACCCGAGG	2562	
Db	721	ACAGAGCAAGGCGCTTGGCAAGACATAGACACACCCAAACCCAAAGGCGTGGACCCGAGG	780	
QY	2563	CCCGGCGCGTGTATCCGAGAGGTGGACATTCAGACTTCCGCTCTGACAGTCCAGGCTC	2622	
Db	781	CCCGGCGCGTGTATCCGAGAGGTGGACATTCAGACTTCCGCTCTGACAGTCCAGGCTC	840	
QY	2623	CTCACACAGAACACCAAGGCGCTGTGTCTCCGAGAGCTTCTTCAAGACCTTCTTCAAGTGC	2682	
Db	841	CTCACACAGAACACCAAGGCGCTGTGTCTCCGAGAGCTTCTTCAAGACCTTCTTCAAGTGC	900	
QY	2683	CCACTTGGAGATGACGAATGACAGCGGAGCTTAGAGACCCCTCCACGCGGCTGGACCTCGGCTG	2742	
Db	901	CCACTTGGAGATGACGAATGACAGCGGAGCTTAGAGACCCCTCCACGCGGCTGGACCTCGGCTG	960	
QY	2743	CAGTAAAGTTACGTGAGGCGCTGTCTCTCGGGGCGCTGGAAATGGGACGCATCAGTGTCT	2802	
Db	961	CAGTAAAGTTACGTGAGGCGCTGTCTCTCGGGGCGCTGGAAATGGGACGCATCAGTGTCT	1020	
QY	2803	TGCTGACCCCTCGAGACGACGCGCGCACAGGTGTGGCTGAGACAGCTGGCGGGGGGCG	2862	
Db	1021	TGCTGACCCCTCGAGACGACGCGCGCACAGGTGTGGCTGAGACAGCTGGCGGGGGGCG	1080	
QY	2863	CCCAAGGCTGGCGCGGCGCTCCAGGCGCACCCACAGCTGTGTGGAAGTCAGGCGCAACTCCCG	2922	
Db	1081	CCCAAGGCTGGCGCGGCGCTCCAGGCGCACCCACAGCTGTGTGGAAGTCAGGCGCAACTCCCG	1140	
QY	2923	CAGCACTGTATCTGAGTAAACGGCTAAGAACTCTCTCTGCTGTGTTGAAAAAGCAATTC	2982	
Db	1141	CAGCACTGTATCTGAGTAAACGGCTAAGAACTCTCTCTCTGCTGTGTTGAAAAAGCAATTC	1200	
QY	2983	GGGTTGTCCAAATCTGTAAACATTCATCCATTTTTT-AAAAAGTTTTCTGTACGACCC	3042	
Db	1201	GGGTTGTCCAAATCTGTAAACATTCATCCATTTTTT-AAAAAGTTTTCTGTACGACCC	1259	
QY	3043	CACGGCGCGAGCGCGGCTGAGCGCTGTGTGATAGACCTGGGCGCCGGGCGCTTCCGCTGC	3102	
Db	1260	CACGGCGCGAGCGCGGCTGAGCGCTGTGTGATAGACCTGGGCGCCGGGCGCTTCCGCTGC	1319	
QY	3103	GCTCTGCGCAGGATGCTTCTGAGGACCCCACTCTGCTGCTTTCATTTGACGTGACCTGTA	3162	
Db	1320	GCTCTGCGCAGGATGCTTCTGAGGACCCCACTCTGCTGCTTTCATTTGACGTGACCTGTA	1379	
QY	3163	CAGAAAGCACTCACCAATTAACCTTTCCTGAAAGCAGAAAAA	3213	
Db	1380	CAGAAAGCACTCACCAATTAACCTTTCCTGAAAGCAGAAAAA	1430	
RESULT 9				
ABS64607 standard; cDNA; 1014 BP.				
XX	ABS64607;			
XX	AC			
XX	DT	15-NOV-2002 (first entry)		

DE Human cDNA probe selectively hybridizing to foetal cell mRNA #22.

XX Human; se; foetal liver myeloid cell; probe; erythroblast;  
KW foetal abnormality; maternal blood; differential display;  
KW chromosomal abnormality; single gene disorder; aneuploidy;  
KW nucleotide triplet expansion disorder; trisomy 13; trisomy 21;  
KW Klinefelter syndrome; spina bifida; sickle cell anaemia; thalasassaemia;  
KW Marfan syndrome; Duchenne muscular dystrophy; cystic fibrosis;  
KW Fragile X-syndrome; Friedreich's ataxia; myotonic dystrophy;  
KW Huntington's disease; sex determination.

XX Homo sapiens.  
XX WO200255985-A2.  
XX 18-JUL-2002.  
XX 01-NOV-2001; 2001WO-US045340.  
XX 15-NOV-2000; 2000US-0248882P.  
XX (HOFF ) ROCHE DIAGNOSTICS CORP.  
XX Schueler PA, Xu H, Foltz L, Wu X, Sha Y, Nagy A, Mahoney WC;  
XX WPI; 2002-619108/66.  
XX P-PSDB; ABG79327.

PT Detecting fetal cells in maternal blood sample, useful for diagnosing an  
PT abnormality in a fetal cell, involves using specific nucleic acid probes  
PT that hybridize to fetal cell associated RNAs.

XX Claim 109; Page 168-169; 215pp; English.

XX The invention relates to detecting foetal cells in maternal blood sample,  
XX involving contacting sample with a first probe or performing expression  
XX analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to  
XX that obtained from mature cells to identify RNA/cDNA species useful as  
XX probe, contacting sample with the probe and identifying if the sample  
XX comprises a cell that comprises mRNA that hybridises to the probe. The  
XX probes are identified by differential display analysis using mature liver  
XX cells and foetal liver myeloid cells of less than 22 weeks of gestation.  
XX Also included are: (1) The probe sequences (appearing as AB564556-  
XX AB564618), an isolated nucleic acid molecule having a sequence which is  
XX at least 90% identical to the probe or its complement, or having a  
XX nucleotide sequence identical to at least 15 consecutive nucleotide  
XX residues of the probe, the encoded proteins from the probe or encoding a  
XX naturally occurring allelic variant or fragment; (2) a non-mammalian host  
XX cell containing the probe and (3) an antibody which selectively binds to  
XX the probe encoded protein. The method is useful for detecting a foetal  
XX cell (such as erythroblast or trophoblast) in a maternal blood sample,  
XX which is useful for diagnosing a chromosomal abnormality, single gene  
XX disorder or nucleotide triplet expansion in the gene, in a foetal cell.  
XX For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
XX syndrome), spina bifida, sickle cell anaemia, a thalasassaemia, Marfan  
XX syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
XX syndrome, Friedreich's ataxia, myotonic dystrophy or Huntington's  
XX disease. The probes are also useful for determination of the sex of a  
XX foetus and for detecting the presence of multiple foetuses at early  
XX stages of pregnancy. The present sequence is a foetal cell specific probe  
XX of the invention

XX Sequence 1014 BP; 275 A; 242 C; 284 G; 213 T; 0 U; 0 Other;  
XX SQ

Query Match 31.3%; Score 1006.6; DB 6; Length 1014;  
Best Local Similarity 99.6%; Pred. No. 5.8e-280;  
Matches 1009; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

11 GAAGATGCGCAGCGCTTGTAGCGAGAGAGCTGACATGAATGAAGCTATTACTGCTGCT 70  
|||  
Db 2 GAAGATGCGCAGCGCTTGTAGCGAGAGAGAGCTGACATGAATGAAGCTATTACTGCTGCT 61  
71 GAACGTGCGCAGGAGGCGCTTCTTGAAGAAGCTGAAGAAGCTGCGAGGCTCTGTAT 130

Db	62	GAACGTCGAGGAGAGGCTCTTCTGAAAGCTGAAAGCTGCTACCGAGGCTCTGAT	121
Qy	131	GCTCTACATCCAGACAAAGACAGAGACCAGAGCTCAAGCTCAGAGCGAACAATT	190
Db	122	GCTCTACATCCAGACAAAGACAGAGACCAGAGCTCAAGCTCAGAGCGAACAATT	189
Qy	191	TAACTTTGTCACGAGCTTATGAAAGTCTTAGAGCCCAACAGAGGCTCTATGA	255
Db	182	TAACTTTGTCACGAGGCTTATGAAAGTCTTAGAGCCCAACAGAGGCTCTATGA	244
Qy	251	TATATATGGGAGAGAGAGCTGGAAATGGAAGATGGGAGGTTGTGAAAGAGAGAAC	311
Db	242	TATATATGGGAGAGAGAGCTGGAAATGGAAGATGGGAGGTTGTGAAAGAGAGAAC	301
Qy	311	CCCTGCTGAAATTCGAGAGAGATTGAGCGGCTGACAGAGAGAGAGAGAGATT	370
Db	302	CCCTGCTGAAATTCGAGAGAGATTGAGCGGCTGACAGAGAGAGAGAGAGATT	361
Qy	371	GCAGACGCGAACCAATCCCAAGGGAACGATCAGCGTTGAGTAGATGACCGACTTTT	430
Db	362	GCAGACGCGAACCAATCCCAAGGGAACGATCAGCGTTGAGTAGATGACCGACTTTT	421
Qy	431	TGATCGCTATGATGAGAGATAGAAAGTGTGTCCGCGAGTAGCTTTCCGACGATTGAAT	490
Db	422	TGATCGCTATGATGAGAGATAGAAAGTGTGTCCGCGAGTAGCTTTCCGACGATTGAAT	481
Qy	491	TAATTAATAATGACATATCCCACTGATCCATTGAGGCACTCTTGAACAGACACAGCAT	550
Db	482	TAATTAATAATGACATATCCCACTGATCCATTGAGGCACTCTTGAACAGACACAGCAT	541
Qy	551	CCTCTCTGGAAGCTCTCAACCAGAAATGGAATGGAAGAGGTTCCATTAACTTTGGCT	610
Db	542	CCTCTCTGGAAGCTCTCAACCAGAAATGGAATGGAAGAGGTTCCATTAACTTTGGCT	601
Qy	611	CAGACGAGTAACTTTCGGTAAAGGATGGGAGAGATTGGAATTTGAGCTGGAGACTTACA	670
Db	602	CAGACGAGTAACTTTCGGTAAAGGATGGGAGAGATTGGAATTTGAGCTGGAGACTTACA	661
Qy	671	GGGAGCCTTTGTTCCGCTCAAGCTGTCCGTAATCTCACCAAGATGCTTTGAGCAAC	730
Db	662	GGGAGCCTTTGTTCCGCTCAAGCTGTCCGTAATCTCACCAAGATGCTTTGAGCAAC	721
Qy	731	AAACTGTGCTGACAGTTTTCATCCCGTGAATCCGACCCGCGCTGACACTGTCTTAGC	790
Db	722	AAACTGTGCTGACAGTTTTCATCCCGTGAATCCGACCCGCGCTGACACTGTCTTAGC	781
Qy	791	TGGAACCTTGAACAGAACACCGTGGGCTACCTGCAATGGCGATTCAGTCAGC	856
Db	782	TGGAACCTTGAACAGAACACCGTGGGCTACCTGCAATGGCGATTCAGTCAGC	841
Qy	851	CATGAACACTAGATGCTCCGAGACATTAACCAAGCACTTCACTGTGGCCCGCAGCT	910
Db	842	CATGAACACTAGATGCTCCGAGACATTAACCAAGCACTTCACTGTGGCCCGCAGCT	901
Qy	911	GGGAATCCCTCACTCTTTGCACTGATAGCATATCAGACAAATTCAGATGACATCA	970
Db	902	GGGAATCCCTCACTCTTTGCACTGATAGCATATCAGACAAATTCAGATGACATCA	961
Qy	971	GACTCGTGAAGAGATCCCTCAAGCAGGCTTTTGAGACGGTGTGAGT 1023	
Db	962	GACTCGTGAAGAGATCCCTCAAGCAGGCTTTTGAGACGGTGTGAGT 1014	
RESULT 10			
AA160714/c			
ID AA160714 standard; cDNA; 1593 BP.			
AA160714;			
AC			
XX 22-Oct-2001 (first entry)			
DE Human polynucleotide SEQ ID NO 4703.			

XX Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200153312-A1.  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-0052317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSEQ-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM41558.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Claim 1; SEQ ID NO 4703; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX  
 SQ Sequence 1593 BP; 336 A; 445 C; 435 G; 375 T; 0 U; 2 Other;

Query Match 25.9%; Score 832.2; DB 4; Length 1593;  
 Best Local Similarity 99.1%; Pred. No. 2.2e-229;  
 Matches 83; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAAAGTTGCGAAGATGCGACGACCTTGTGAGCAGAGAGCTGCAATGAAGACTATT 60  
 DB 1455 GAAAGTTGCGAAGATGCGACGACCTTGTGAGCAGAGAGCTGCAATGAAGACTATT 1396

QY 61 ACTGTTTCTGAAAGTGCAGAGGAGCTTTCTGAAGAGCTGAAAGCTGCTTACCGGA 120  
 DB 1395 ACTGTTTCTGAAAGTGCAGAGGAGCTTTCTGAAGAGCTGAAAGCTGCTTACCGGA 1336

QY 121 GGGCTGTATGCTCTACCAATCCAGACAGACAGACAGACCCAGAGCTCAAGTCAAGCGG 180  
 DB 1335 GGGCTGTATGCTCTACCAATCCAGACAGACAGACAGACCCAGAGCTCAAGTCAAGCGG 1276

QY 181 AACGACTTTTAACCTTTGTTACACAGGCTTATGAAGTGTATGACCCCAACAGGG 240  
 DB 1275 AAGCACTGTTTAACCTTTGTTACACAGGCTTATGAAGTGTATGACCCCAACAGGG 1216

QY 241 CCATCTATATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGTTGTGAAA 300  
 DB 1215 CCATCTATATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGTTGTGAAA 1156

QY 301 GGAGGAGAACCCCTGCTGAATTCGAGAGAGATTTGAGGCGTGCAGAGAGAGAGAG 360  
 DB 1155 GGAGGAGAACCCCTGCTGAATTCGAGAGAGATTTGAGGCGTGCAGAGAGAGAGAG 1096

QY 361 AGAGAGATTTGAG 420  
 DB 1095 AGAGAGATTTGAG 1036

QY 421 CCGACCTTTTGTATGCTATATGATGAGAGATATGAAGATGTGTCGGAGTAGCTTCCGC 480  
 DB 1035 CCGACCTTTTGTATGCTATGATGAGAGATATGAAGATGTGTCGGAGTAGCTTCCGC 976

QY 481 AGATTGAATTAATAATGCAATATCCAGTCCATTGAGGACCCCTTGACAGAGAGAG 540  
 DB 975 AGATTGAATTAATAATGCAATATCCAGTCCATTGAGGACCCCTTGACAGAGAGAG 916

QY 541 ACACAGCCATCTCTCTGGAAGCTCTCAACCCAGAGATGAATGAGAGAGGTTCCATT 600  
 DB 915 ACACAGCCATCTCTCTGGAAGCTCTCAACCCAGAGATGAATGAGAGAGGTTCCATT 856

QY 601 ACTTTGCGCTCAGACGAGATTAACCTTGGCAAGAGATGGGAGAGTTGGAATTTGAAGCTG 660  
 DB 855 ACTTTGCGCTCAGACGAGATTAACCTTGGCAAGAGATGGGAGAGTTGGAATTTGAAGCTG 796

QY 661 GAGACCTAAGGGGCTTTGTTGGTTCGATCTCAACCTGTCGATCTCAACCAAGATGCT 720  
 DB 795 GAGACCTAAGGGGCTTTGTTGGTTCGATCTCAACCTGTCGATCTCAACCAAGATGCT 736

QY 721 TTGTGACAAACAACTGTGCTGAGATTTTCATCCGTGGAATCCGACCCGAGCTGACCA 780  
 DB 735 TTGTGACAAACAACTGTGCTGAGATTTTCATCCGTGGAATCCGACCCGAGCTGACCA 676

QY 781 CTGTCTTACCTGGAACCTTGAACAGAGACAGAGAGAGAGAGAGAGAGAGAGAG 840  
 DB 675 CTGTCTTACCTGGAACCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 616

QY 841 TCCAG 845  
 DB 615 CCCCC 611

RESULT 11  
 AA158928  
 ID AA158928 standard; cDNA; 1767 BP.  
 XX  
 AC AA158928;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 1131.  
 XX  
 KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200153312-A1.  
 PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-0048725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
 PI Wang Z, Wehran T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR P-ESDB: AAM39772.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Claim 1; SEQ ID NO 1131; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AAM38642-AA42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 CC  
 XX  
 SQ Sequence 1767 BP; 393 A; 503 C; 493 G; 376 T; 0 U; 2 Other;  
 Query Match 25.9%; Score 832.2; DB 4; Length 1767;  
 Best Local Similarity 99.1%; Pred. No. 2.3e-229;  
 Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 GAAAGGTTGCGAAGATGGCGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 60  
 DB 139 GAAAGGTTGCGAAGATGGCGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 198  
 QY 61 ACTGCTTGAAGAGCTGCGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 120  
 DB 199 ACTGCTTGAAGAGCTGCGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 258  
 QY 121 GGGCTGTATGCTTACCATCCAGCAAGCAAGAGCCCAAGCTCAAGTCAAGCGG 180  
 DB 259 GGGCTGTATGCTTACCATCCAGCAAGCAAGAGCCCAAGCTCAAGTCAAGCGG 318  
 QY 181 AAGAGCTTTAACTGTTACACAGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 240  
 DB 319 AAGAGCTTTAACTGTTACACAGGCTTGAAGAGCTGAAAGCTGCTTACCGGA 378  
 QY 241 CCATCTATGATATATATGGAAGAGAGCTGAAATGGAAGATGGAGGTTTGAAA 300  
 DB 379 CCATCTATGATATATATGGAAGAGAGCTGAAATGGAAGATGGAGGTTTGAAA 438  
 QY 301 GGAGGAGAAACCCCTGCTGAAATTTGAGAGAGCTTGAAGCGGCTGAGAGAGAGAG 360  
 DB 439 GGAGGAGAAACCCCTGCTGAAATTTGAGAGAGCTTGAAGCGGCTGAGAGAGAGAG 498  
 QY 361 AGAGGAGATTGACAGAGAAACAATCCCAAGGAGAGATGAGGTTGAGTATATGCA 420

DB 499 AGAGGAGATTGACAGAGCAACCAATCCCAAGGAGAGATGAGTATGATGCCA 558  
 QY 421 CCGAAGCTTTTGAATGCTATGATGAGAGATGAGATGATGCTCCGAGATGCTTCGCG 480  
 DB 559 CCGAAGCTTTTGAATGCTATGATGAGAGATGAGATGATGCTCCGAGATGCTTCGCG 618  
 QY 481 AGATTGAATTAATTAATTAATGCAATATCCAGTCCATTGAGGCACTTGAACGCGCAG 540  
 DB 619 AGATTGAATTAATTAATTAATGCAATATCCAGTCCATTGAGGCACTTGAACGCGCAG 678  
 QY 541 ACAGAGCATCTCTCTGGAAGCTCTCAACCCAGAAATGGAATGAGAGGTTTCCATTA 600  
 DB 679 ACAGAGCATCTCTCTGGAAGCTCTCAACCCAGAAATGGAATGAGAGGTTTCCATTA 738  
 QY 601 ACTTGGGCTCAGACAGATTAATCTGGCAAGGATGGGAGATTTGGAATTTGAGCTG 660  
 DB 739 ACTTGGGCTCAGACAGATTAATCTGGCAAGGATGGGAGATTTGGAATTTGAGCTG 798  
 QY 661 GAGACCTACAGGGGCTTTTGTGCTCAAGCTGTTCCGTATCTTCAACCAAGATGCT 720  
 DB 799 GAGACCTACAGGGGCTTTTGTGCTCAAGCTGTTCCGTATCTTCAACCAAGATGCT 858  
 QY 721 TTGTGACAAACAATGCTGCTCAAGTTCATCCGTTGGAATCCGACCCGCTGACCA 780  
 DB 859 TTGTGACAAACAATGCTGCTCAAGTTCATCCGTTGGAATCCGACCCGCTGACCA 918  
 QY 781 CTGTCTTAAGTCTGGAACCTTGAACAAAGACCGTGGCTTACCTGCAAGTGGAGTGA 840  
 DB 919 CTGTCTTAAGTCTGGAACCTTGAACAAAGACCGTGGCTTACCTGCAAGTGGAGTGA 978  
 QY 841 TCCAG 845  
 DB 979 CCCC 983  
 RESULT 12  
 ADB48910  
 ID ADB48910 standard; cDNA; 1767 BP.  
 AC ADB48910;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Novel human cDNA SEQ ID NO 820.  
 KW sf; cancer; neurodegenerative disease; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US2003104529-A1.  
 XX  
 PD 05-JUN-2003.  
 PF 04-JAN-2002; 2002US-00037270.  
 XX  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 19-JUL-2000; 2000US-00620312.  
 XX  
 PA (ZHOU/) ZHOU P.  
 PA (TANG/) TANG Y T.  
 PA (LIU/) LIU C.  
 PA (ASUN/) ASUNDI V.  
 PA (DRMA/) DRMANAC R T.  
 XX  
 PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;  
 DR WPI: 2003-678194/64.  
 DR  
 PT New polynucleotide, useful for treating diseases e.g., cancer or  
 PT neurodegenerative diseases.  
 XX



PS Claim 1; SEQ ID NO 820; 99bp; English.  
XX The invention relates to a polynucleotide comprising a sequence given in  
CC the specification, or its mature protein-coding portion, or its  
CC complement. The polynucleotide is useful for treating diseases e.g.,  
CC cancer or neurodegenerative diseases and many others listed in the  
CC specification. The present sequence represents a novel human cDNA. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html?docid=20030104529](http://seqdata.uspto.gov/sequence.html?docid=20030104529).  
XX  
SQ Sequence 1767 BP; 393 A; 503 C; 493 G; 376 T; 0 U; 2 Other;  
  
Query Match 25.9%; Score 832.2; DB 8; Length 1767;  
Best Local Similarity 99.1%; Pred. No. 2.3e-229;  
Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 GAAAGCTTGGGAAAGATGCGACGCGCTTGAAGCGAGAGAGCTGACATGAAGACTATT 60  
DB |||||  
139 GAAAGCTTGGGAAAGATGCGACGCGCTTGAAGCGAGAGAGCTGACATGAAGACTATT 198  
  
QY 61 ACTGCTGTGAAAGTGGGAGAGGAGCTCTTCTGAAGAGCTGAAGAGCTGACCGGAG 120  
DB |||||  
199 ACTGCTGTGAAAGTGGGAGAGGAGCTCTTCTGAAGAGCTGACCGGAG 258  
  
QY 121 GAGCTGTATGCTCTACATCCAGACAGACAGAGAGCCAGAGCTCAAGTCAAGGCGG 180  
DB |||||  
259 GAGCTGTATGCTCTACATCCAGACAGACAGAGAGCCAGAGCTCAAGTCAAGGCGG 318  
  
QY 181 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAGTGTGATGACCCCAACAGAGG 240  
DB |||||  
319 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAGTGTGATGACCCCAACAGAGG 378  
  
QY 241 CCATCTATGATATATATGAGAGAGAGAGCTGGAATGGAAGATGGAGAGTGTGGAAA 300  
DB |||||  
379 CCATCTATGATATATATGAGAGAGAGAGCTGGAATGGAAGATGGAGAGTGTGGAAA 438  
  
QY 301 GAGAGGAAACCCCTGTGTAATTCAGAGAGAGTGTGAGCGCTGACAGAGAGAGAGAG 360  
DB |||||  
439 GAGAGGAAACCCCTGTGTAATTCAGAGAGAGTGTGAGCGCTGACAGAGAGAGAGAG 498  
  
QY 361 AAGAGAGATGTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB |||||  
499 AAGAGAGATGTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558  
  
QY 421 CCGACCTTTTGAATGCTATGATGAGAGAGATGAGAGAGTGTCCGAGATGAGCTTCCG 480  
DB |||||  
559 CCGACCTTTTGAATGCTATGATGAGAGAGATGAGAGAGTGTCCGAGATGAGCTTCCG 618  
  
QY 481 AGATTGAATTAATAAATGACATATCCAGTTCATTGAGGACCCCTTGACAGAGAG 540  
DB |||||  
619 AGATTGAATTAATAAATGACATATCCAGTTCATTGAGGACCCCTTGACAGAGAG 678  
  
QY 541 ACAGAGCATCCCTCTGGAAGAGCTCTCAACCCAGATGGAATGAGAGAGTTCATTA 600  
DB |||||  
679 ACAGAGCATCCCTCTGGAAGAGCTCTCAACCCAGATGGAATGAGAGAGTTCATTA 738  
  
QY 601 ACTTGGGCTCAGACAGATTAATTCGCAAGAGAGAGAGAGAGTGTGAATTTGAGAGT 660  
DB |||||  
739 ACTTGGGCTCAGACAGATTAATTCGCAAGAGAGAGAGAGTGTGAATTTGAGAGT 798  
  
QY 661 GAGACCTACAGGGGAGCTTTGCTGAGTCAAGTGTTCCTTAATCTCACACCAAGATGCT 720  
DB |||||  
799 GAGACCTACAGGGGAGCTTTGCTGAGTCAAGTGTTCCTTAATCTCACACCAAGATGCT 858  
  
QY 721 TTGTGACAAACAACTGTGCTCTGAGATTTCAATCCCGTGAATCCAGCCGAGCTGACCA 780  
DB |||||  
859 TTGTGACAAACAACTGTGCTCTGAGATTTCAATCCCGTGAATCCAGCCGAGCTGACCA 918  
  
QY 781 CTGTCTAGAGTGGAACTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
DB |||||  
919 CTGTCTAGAGTGGAACTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 978

QY 841 TCCAG 845  
DB |||  
979 CCCCC 983  
  
RESULT 13  
AAS78884  
ID AAS78884 standard; cDNA; 1095 BP.  
XX  
XX AAS78884;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX DNA encoding novel human diagnostic protein #14688.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR  
XX  
XX 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSBQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX P-PSDB; ABB14697.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX  
XX Claim 1; SEQ ID NO 14688; 103bp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probe, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
XX  
SQ Sequence 1095 BP; 270 A; 281 C; 319 G; 225 T; 0 U; 0 Other;  
  
Query Match 25.5%; Score 819.6; DB 5; Length 1095;  
Best Local Similarity 98.8%; Pred. No. 7.7e-226;  
Matches 836; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
  
QY 1 GAAAGCTTGGGAAAGATGCGACGCGCTTGAAGCGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB |||||

```
Db 13 GAAAGTTGCGAAGATGCGCAGCGCTTGAAGCAGAGAGAGCTGGACAATGAGACTATT 72
Qy 61 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 120
Db 73 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 132
Qy 121 GGCTCTGATGCTCTACCATCAGACAGACAGAGAGCCGAGAGCTCAAGTCAAGGCGG 180
Db 133 GGCTCTGATGCTCTACCATCAGACAGACAGAGAGCCGAGAGCTCAAGTCAAGGCGG 192
Qy 181 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 240
Db 193 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 252
Qy 241 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 300
Db 253 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 312
Qy 301 GGAGGAGAACCCCTGCG-TGAAATTGAGAGAGAGTTGAGCGGCTGAGAGAGAGAGAA 359
Db 313 GGAGGAGAACCCCTGCGTGGAAATTCGAGAGAGAGTTGAGCGGCTGAGAGAGAGAA 372
Qy 360 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 373 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
Qy 420 ACCGACCTTTTGTATCGCTATGATGAGAGAGATGAAAGTGTCCGAGTATGCTTCCG 479
Db 433 ACCGACCTTTTGTATCGCTATGATGAGAGAGATGAAAGTGTCCGAGTATGCTTCCG 492
Qy 480 CAGATTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 539
Db 493 CAGATTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
Qy 540 GACACAGCCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
Db 553 GACACAGCCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
Qy 600 AACTTTGGGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
Db 613 AACTTTGGGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
Qy 660 GAGAGACCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Db 673 GAGAGACCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
Qy 720 TTTGTGACAAACAACTGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
Db 733 TTTGTGACAAACAACTGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792
Qy 780 ACTGTCTAGCTCGAAGCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
Db 793 ACTGTCTAGCTCGAAGCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 852
Qy 840 ATCCAG 845
Db 853 TCCCCG 858
```

## RESULT 14

AAS67386 standard; cDNA, 1098 BP.

AAS67386;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #3190.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.  
Homo sapiens.

```
XX 13 GAAAGTTGCGAAGATGCGCAGCGCTTGAAGCAGAGAGAGCTGGACAATGAGACTATT 72
PN 13 GAAAGTTGCGAAGATGCGCAGCGCTTGAAGCAGAGAGAGCTGGACAATGAGACTATT 72
XX 61 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 120
PD 61 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 132
XX 73 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 132
PF 73 ACTGTTGCTGAAGAGTGGCGAGAGGCTCTTCTGAAGAGCTGAAGAGCTCCATCCGGA 132
XX 121 GGCTCTGATGCTCTACCATCAGACAGACAGAGAGCCGAGAGCTCAAGTCAAGGCGG 180
PR 121 GGCTCTGATGCTCTACCATCAGACAGACAGAGAGCCGAGAGCTCAAGTCAAGGCGG 192
XX 181 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 240
PA 181 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 252
XX 193 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 252
PI 193 AACGACTGTTAACTTGTTCACCAAGGCTTATGAAAGTCTTATGAGACCCCAACAGGG 252
XX 241 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 300
PT 241 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 312
XX 253 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 312
PT 253 CCATCTATGATATATATATGAGAGAGAGAGCTGGAAATGGAAGATGGAGGTTGTGAAA 312
XX 301 GGAGGAGAACCCCTGCG-TGAAATTGAGAGAGAGTTGAGCGGCTGAGAGAGAGAGAA 359
PT 301 GGAGGAGAACCCCTGCG-TGAAATTGAGAGAGAGTTGAGCGGCTGAGAGAGAGAGAA 359
XX 313 GGAGGAGAACCCCTGCGTGGAAATTCGAGAGAGAGTTGAGCGGCTGAGAGAGAGAA 372
PT 313 GGAGGAGAACCCCTGCGTGGAAATTCGAGAGAGAGTTGAGCGGCTGAGAGAGAGAA 372
XX 360 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
PS 360 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
XX 373 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
PS 373 GAGAGAGATTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
XX 420 ACCGACCTTTTGTATCGCTATGATGAGAGAGATGAAAGTGTCCGAGTATGCTTCCG 479
XX 433 ACCGACCTTTTGTATCGCTATGATGAGAGAGATGAAAGTGTCCGAGTATGCTTCCG 492
XX 480 CAGATTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 539
XX 493 CAGATTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
XX 540 GACACAGCCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
XX 553 GACACAGCCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
XX 600 AACTTTGGGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
XX 613 AACTTTGGGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
XX 660 GAGAGACCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
XX 673 GAGAGACCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
XX 720 TTTGTGACAAACAACTGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
XX 733 TTTGTGACAAACAACTGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792
XX 780 ACTGTCTAGCTCGAAGCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
XX 793 ACTGTCTAGCTCGAAGCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 852
XX 840 ATCCAG 845
XX 853 TCCCCG 858
```

Query Match 25.5%; Score 819.6; DB 5; Length 1098;  
Best Local Similarity 98.8%; Pred. No. 7.7e-226;  
Matches 836; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Sequence 1098 BP; 271 A; 281 C; 320 G; 226 T; 0 U; 0 Other;

Claim 1; SEQ ID NO 3190; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Query Match 25.5%; Score 819.6; DB 5; Length 1098;  
Best Local Similarity 98.8%; Pred. No. 7.7e-226;  
Matches 836; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Sequence 1098 BP; 271 A; 281 C; 320 G; 226 T; 0 U; 0 Other;

Claim 1; SEQ ID NO 3190; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences



Db	313	GGAGGGAACCCCTGCTTGAATTCGAGAGAGTTTGAGCGGCTCTGAGGAGAGGAA	3722
Qy	360	GAGAGGAGATTGCGAGCGGAAACCAATCCAGGGAGCAATCAGCGTTGAGTAGATGCC	419
Db	373	GAGAGGAGATTGCGAGCGGAAACCAATCCAGGGAGCAATCAGCGTTGAGTAGATGCC	4322
Qy	420	ACCGACCTTTTGTATGCTTATGATGAGAGATATGAAGATGTGTCCGGCAGTAGCTTTCCG	479
Db	433	ACCGACCTTTTGTATGCTTATGATGAGAGATATGAAGATGTGTCCGGCAGTAGCTTTCCG	492
Qy	480	CAGATTGAAATTAATAAATGACATATCCAGTCAATTGAGGACCCCTTGACAGGACA	539
Db	493	CAGATTGAAATTAATAAATGACATATCCAGTCAATTGAGGACCCCTTGACAGGACA	552
Qy	540	GACACAGCCATCCCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGTTCCATT	599
Db	553	GACACAGCCATCCCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGTTCCATT	612
Qy	600	AACCTTGGCGCTGACAGAGTAACCTTGGCAAGGAGTAGGGAGAGTTGGAGCT	659
Db	613	AACCTTGGCGCTGACAGAGTAACCTTGGCAAGGAGTAGGGAGAGTTGGAGCT	672
Qy	660	GGAGACCTACAGGGGCGCTTGTTCGGTCAAGCTGTTCCGTAATCTCACACCAAGATGC	719
Db	673	GGAGACCTACAGGGGCGCTTGTTCGGTCAAGCTGTTCCGTAATCTCACACCAAGATGC	732
Qy	720	TTTGTGCACAAAACTGTGCTCTGACAGTTTTCATCCGTGGAAATCCGACCCGGCCTGACC	779
Db	733	TTTGTGCACAAAACTGTGCTCTGACAGTTTTCATCCGTGGAAATCCGACCCGGCCTGACC	792
Qy	780	ACTGTCTCTAGCTCGAACTTACGCAAGAACACCGTGGCTTACTGTGACGTGGAGTGGGT	839
Db	793	ACTGTCTCTAGCTCGAACTTACGCAAGAACACCGTGGCTTACTGTGACGTGGAGTGGGT	852
Qy	840	ATCCAG 845	
Db	853	TCCCCG 858	
RESULT 15			
AAH06495			
ID	AAH06495	standard; cDNA; 706 BP.	
XX	AAH06495;		
AC			
DT	26-JUN-2001	(first entry)	
XX			
DE	Human cDNA clone (5'-primer) SEQ ID NO:3330.		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	Homo sapiens.		
OS			
XX	EPI074617-A2.		
PN			
XX	07-FEB-2001.		
PD			
XX			
PF	28-JUL-2000; 2000EP-00116126.		
XX			
PR	29-JUL-1999; 99JP-00248036.		
XX			
PR	27-AUG-1999; 99JP-00300253.		
XX			
PR	11-JAN-2000; 2000JP-00118776.		
XX			
PR	02-MAY-2000; 2000JP-00183767.		
XX			
PR	09-JUN-2000; 2000JP-00241899.		
XX			
PA	(HELI-) HELIX RES INST.		
PI	Oca T, Isogai T, Nishikawa T, Hayaeshi K, Saito K, Yamamoto J;		
PI	Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
XX	WPI; 2001-318749/34.		

PT		Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and/or
PT		diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX		
PS		Claim 1; SEQ ID NO 3330; 2537bp + Sequence Listing; English.
XX		
CC	The present invention describes primer sets for synthesizing 5602 full-	
CC	length cDNAs defined in the specification. Where a primer set comprises:	
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the	
CC	complementary strand of a polynucleotide which comprises one of the 5602	
CC	nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the	
CC	specification. The primer sets can be used in antisense therapy and in	
CC	gene therapy. The primers are useful for synthesising polynucleotides,	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893	
CC	represent human amino acid sequences; and AAH13629 to AAH13633 represent	
CC	oligonucleotides, all of which are used in the exemplification of the	
CC	present invention	
XX		
SQ	Sequence 706 BP; 199 A; 151 C; 208 G; 144 T; 0 U; 4 Other;	
	Query Match 21.3%; Score 684.2; DB 4; Length 706;	
	Best Local Similarity 98.9%; Pred. No. 8.8e-187;	
	Matches 697; Conservative 0; Mismatches 7; Indels 1; Gaps 1	
OY	2 AAAAGTTGGGAAGATGCGACGGCCTTGAGCGAAGACGTGCACAATTAGAATACTATTA 61	
Dd	3 AAAAGTTGGGAAGATGCGACGGCCTTGAGCGAAGACGTGCACAATTAGAATACTATTA 62	
OY	62 CTCGTTGCCGAACGTCGCAAGGAGGCGCTCTTGAAAGACTGAAAAGCTGCCACCGGAG 121	
Dd	63 CTCGTTGCCGAACGTCGCAAGGAGGCGCTCTTGAAAGACTGAAAAGCTGCCACCGGAG 122	
OY	122 GCCTGTATGCTCTACCATCCAGACAAGCACAGAGACCAGAGCTCAAGTCAcAGCGGGA 181	
Dd	123 GCCTGTATGCTCTACCATCCAGACAAGCACAGAGACCAGAGCTCAAGTCAcAGCGGGA 182	
OY	182 AGCACTGTTAACCTGGTTCACACAGGCTTATGAAGTGCTTAGTACCCTCCAACCGAGGC 241	
Dd	183 AGCACTGTTAACCTGGTTCACACAGGCTTATGAAGTGCTTAGTACCCTCCAACCGAGGC 242	
OY	242 CATCTATGATAATATATGGAAGAGAGACTGGAAAAATGGAAGATGGAGATTGTGGAAG 301	
Dd	243 CATCTATGATAATATATGGAAGAGAGACTGGAAAAATGGAAGATGGAGATTGTGGAAG 302	
OY	302 GAGGAGAAACCCCTGCTGAAATTTGAGAGAGAGTTTGACGGCTGCAGAGAGAGAGAA 361	
Dd	303 GAGGAGAAACCCCTGCTGAAATTTGAGAGAGAGTTTGACGGCTGCAGAGAGAGAGAA 362	
OY	362 GAGGAGATTGcAGAcGAGCAAATCCCAGAGGAACGATCAGCGTTGGAGTAGATGCCNC 421	
Dd	363 GAGGAGATTGcAGAcGAGCAAATCCCAGAGGAACGATCAGCGTTGGAGTAGATGCCAC 422	
OY	422 CGACCTTTTTGAATCCCTATGATGAGAGATATGAAGATGTGTCCGGCAGTATGCTTTCCGA 481	
Dd	423 CGACCTTTTTGAATCCCTATGATGAGAGATATGAAGATGTGTCCGGCAGTATGCTTTCCGA 482	
OY	482 GATTGAAATTAATAAAATGCAATATCCCAATGCATATGAGAGACCTTGACAGCGACAGA 541	
Dd	483 GATTGAAATTAATAAAATGCAATATCCCAATGCATATGAGAGACCTTGACAGCGACAGA 542	
OY	542 CAcAGcATccCTctTGAAgcCTtCcAAcCCAgAtgAAATGAgAGgtTCcATTAA 601	



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:52:48 ; Search time 218 Seconds  
(Without alignment)

8179.162 Million cell updates/sec

Title: US-09-787-678a-12

Perfect score: 3213

Sequence: 1 gaaagctgcgaagatgctg.....gaaagcagaaaaaaaaa 3213

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patente NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832.2	25.9	1767	4	US-09-620-312D-820
2	455	14.2	455	4	US-09-404-879A-326
3	83.2	2.6	1261	4	US-09-724-864-34
4	58	1.8	1330	2	US-08-868-288A-4
5	58	1.8	1330	3	US-09-235-373-4
6	58	1.8	1330	3	US-09-388-993-4
7	57.2	1.8	2349	2	US-08-974-546-2
8	56.8	1.8	7218	1	US-08-233-463-14
9	48.4	1.5	1308	4	US-09-023-655-883
10	48.4	1.5	1318	4	US-09-665-479A-7
11	48.4	1.5	2607	4	US-09-976-594-226
12	44.2	1.4	713	4	US-09-023-655-300
13	42	1.3	297	4	US-09-313-294A-6215
14	42	1.3	1700	2	US-08-897-340-4
15	42	1.3	1700	3	US-09-252-329-4
16	42	1.3	3748	2	US-08-686-417-2
17	41	1.3	505	4	US-09-621-976-15639
18	40.4	1.3	284	4	US-09-313-294A-1936
19	40.4	1.3	1756	2	US-08-879-260-3
20	40.4	1.3	1811	3	US-09-231-529-5
21	40.4	1.3	1811	3	US-08-977-816-5
22	40.2	1.3	635	4	US-09-669-751-166
23	40.2	1.3	4403765	3	US-09-103-840A-2
24	40.2	1.3	4411529	3	US-09-103-840A-1
25	39.2	1.2	1166	4	US-09-072-596-323
26	39.2	1.2	1166	4	US-09-072-967-328
27	38.8	1.2	364	4	US-09-621-976-17202

28	38.6	1.2	1482	4	US-09-613-303-18	Sequence 18, Appl
29	38.6	1.2	1482	4	US-10-267-311-18	Sequence 18, Appl
30	38.6	1.2	16442	3	US-08-781-891-208	Sequence 208, App
31	38.6	1.2	16442	4	US-09-618-166-208	Sequence 208, App
32	38.2	1.2	1926	4	US-09-249-585A-4	Sequence 4, Appl1
33	38.2	1.2	1931	2	US-09-130-114-2	Sequence 2, Appl1
34	38.2	1.2	2019	3	US-09-063-950-3	Sequence 3, Appl1
35	38.2	1.2	2555	4	US-09-866-028-68	Sequence 68, Appl
36	38.2	1.2	2852	3	US-09-063-950-1	Sequence 1, Appl1
37	38.2	1.2	6136	4	US-08-956-171E-219	Sequence 219, App
38	37.8	1.2	882	4	US-09-252-991A-2427	Sequence 2427, App
39	37.8	1.2	993	4	US-09-252-991A-2568	Sequence 2568, App
40	37.6	1.2	2481	4	US-09-894-998A-35	Sequence 35, Appl
41	37.6	1.2	5919	4	US-08-875-435B-2	Sequence 2, Appl1
42	37.4	1.2	505	4	US-09-621-976-15639	Sequence 15639, A
43	37.4	1.2	1755	1	US-08-068-195A-2	Sequence 2, Appl1
44	37.4	1.2	1755	1	US-08-464-365-2	Sequence 2, Appl1
45	37.4	1.2	2568	4	US-09-489-039A-5258	Sequence 5258, Ap

## ALIGNMENTS

RESULT 1  
US-09-620-312D-820

Sequence 820, Application US/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Zhao, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yundong

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6569662el Nucleic Acids and

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/09/620,312D

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: pc\_fl\_genes Version 1.0

SEQ ID NO 820

LENGTH: 1767

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (153)..(1223)

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)..(1767)

OTHER INFORMATION: n = a,t,c or g

US-09-620-312D-820

Query Match 25.9%; Score 832.2; DB 4;

Best Local Similarity 99.1%; Pred. No. 1,5e-227;

Matches 837; Conservative 8; Indels 0; Gaps 0;

```
QY      1 GAAAGTTGGGAATGCGACGCGCTTGAGCGAGGAGCTGACAAATGAAGCTATT 60
      139 GAAGGTGGGAATGCGACGCGCTTGAGCGAGGAGCTGACAAATGAAGCTATT 198
QY      61 ACTGTTGCTGAACTGTCGCGAGGAGGCTCTTCTGAGAGCTGAAAGCTGCTTACCGA 120
      199 ACTGTTGCTGAACTGTCGCGAGGAGGCTCTTCTGAGAGCTGAAAGCTGCTTACCGA 258
QY      121 GGGCTGTATGCTCTACCAATCCAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 180
      259 GGGCTGTATGCTCTACCAATCCAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 318
QY      181 AAGCACTGTTAACTGTTTCAACAGGCTTATGAGTGTCTTATGACCCCAACCAAGG 240
      319 AAGCACTGTTAACTGTTTCAACAGGCTTATGAGTGTCTTATGACCCCAACCAAGG 378
QY      241 CCACTATGATATATATGAGAGAGAGACTGGAATGGAAGATGGAGGTTTGAGAA 300
      379 CCACTATGATATATATGAGAGAGAGACTGGAATGGAAGATGGAGGTTTGAGAA 438
QY      301 GGAAGAGACCCCTGCTGAAATTCAGAGAGGTTTGAGCGGCTGACAGAGAGAGAG 360
      439 GGAAGAGACCCCTGCTGAAATTCAGAGAGGTTTGAGCGGCTGACAGAGAGAGAG 498
QY      439 GGAAGAGACCCCTGCTGAAATTCAGAGAGGTTTGAGCGGCTGACAGAGAGAGAG 498
      499 AGAGAGATTTGACAGCGACCAATCCCAAGGAGAGATCAAGCTTGGATGATGCCA 558
QY      421 CCGAAGCTTTTGAATCGCTATGATGAGAGATGATGATGTCGCGAGTATGCTTCCG 480
      559 CCGAAGCTTTTGAATCGCTATGATGAGAGATGATGATGTCGCGAGTATGCTTCCG 618
QY      481 AGATGAAATTAATAAATGCAATATCCAGTCCATGAGGAGACCTTTGACAGCGACG 540
      619 AGATGAAATTAATAAATGCAATATCCAGTCCATGAGGAGACCTTTGACAGCGACG 678
QY      541 ACAAGCCATCTCTCTGAGAGCTCTCAACCCAGATGAGAAATGAGAGGTTTCCATT 600
      679 ACAAGCCATCTCTCTGAGAGCTCTCAACCCAGATGAGAAATGAGAGGTTTCCATT 738
QY      601 ACTTTGGCTGACAGAGTAACTTCCGCAAGAGATGGAGAGTTGGAATTTGAGCTG 660
      739 ACTTTGGCTGACAGAGTAACTTCCGCAAGAGATGGAGAGTTGGAATTTGAGCTG 798
QY      661 GAGACCTACAGGGGCTTTGTTGGTCTCAAGCTGTTCCGTAATCTCACCAAGATGCT 720
      799 GAGACCTACAGGGGCTTTGTTGGTCTCAAGCTGTTCCGTAATCTCACCAAGATGCT 858
QY      721 TTGTGACAAACAACTGTCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGCTGAC 780
      859 TTGTGACAAACAACTGTGTCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGCTGAC 918
QY      781 CTGTCTTACGTCGGAACCTTGAACAACACCGTGGCTTCTGCAAGTGGCGATG 840
      919 CTGTCTTACGTCGGAACCTTGAACAACACCGTGGCTTCTGCAAGTGGCGATG 978
QY      841 TCCAG 845
      979 CCCCC 983
Db
```

```
RESULT 2
US-09-404-879A-326/c
; Sequence 326, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
```

```
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 326
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-404-879A-326
```

```
Query Match      14.2%; Score 455; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 5,1e-120; Indels 0; Gaps 0;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1074 AGCGTTGAGTTCCACAGGCGCTTTCTCTCAAGTCAAGCTCAACAGGCGCAGTACAGACA 1133
      455 AGCGTTGAGTTCCACAGGCGCTTTCTCTCAAGTCAAGCTCAACAGGCGCAGTACAGACA 396
QY      1134 TACTTCTTCCATTACCTTGAACGACAGCTTTCGCCAGCGCATGTTCTATGCGACC 1193
      395 TACTTCTTCCATTACCTTGAACGACAGCTTTCGCCAGCGCATGTTCTATGCGACC 336
QY      1194 GTGGGCGCTCTAGTGTCTACTTTTGCATGACACCGTGTATATCAACACATACCTCAGG 1253
      335 GTGGGCGCTCTAGTGTCTACTTTTGCATGACACCGTGTATATCAACACATACCTCAGG 276
QY      1254 GCTCAAGAAAGAAAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
      275 GCTCAAGAAAGAAAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 216
QY      1314 AAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373
      215 AAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 156
QY      1374 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1433
      155 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 96
QY      1434 AATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1493
      95 AATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 36
QY      1494 TGCCTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1528
      35 TGCCTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1
Db
```

```
RESULT 3
US-09-724-864-34
; Sequence 34, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; FILE REFERENCE: 11000.1050U1
; CURRENT APPLICATION NUMBER: US/09/724,864
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1261
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-864-34
```

```
Query Match      2.6%; Score 83.2; DB 4; Length 1261;
Best Local Similarity 55.8%; Pred. No. 2e-13; Indels 72; Gaps 10;
Matches 404; Conservative 0; Mismatches 248; Indels 72; Gaps 10;
QY      1703 GAACGAGATTTTAAAGGCGCAAAAATCTTTCTGCGAGTCTACAAATTTGGAA-A 1761
```

Db	24	GAACCTCGAGGAAAAAGAGCGTACGAAAAA	CTTTTCTTAAAGCGCTACAAATTGGAAAC	83	
Oy	1762	TGAAAAAA	CCGACATCATGATGTTTTATTTTTATTA	TTATTAATTATTAAGAGGTGTACCA	1822
Db	84	AGGGAAAA	CCCGACATGATGTTTTAGTTATTCTCCAGAA	-----GGGGGCACT	137
Oy	1822	TTATCAATTATGTGAAGGACA	TGAGACA	CCCCAGCTTTGAGGGTGTCTGGGGGTAGGA	1883
Db	138	GTATCAATTATGTGAAGGACA	TGAGACA	CAAGCTTGACCTGCATGTGTGTGT-GGGGGTAGGA	196
Oy	1882	CTGAGGACGCC	CCCACTGGGAACCA	CACTGCAGCTGGCCCATGCTGTTTTCCCAAGAT	1941
Db	197	CTGAGGACCC	CTGTGGCCAGACCCAGACGGCCATG	-----TCTCTCCAAAGATC	248
Oy	1942	CAGTTCTTGAAGGAAAGG	CTCTGAGCCCTGACTCCGCTGTGTCCGAGACACGTGCTGA	2001	
Db	249	ATGTTCTTGAAGGTAC	CGCCCTGTGCTT-----	278	
Oy	2002	CCGAGACCCG	CGCCCTGTGATCTTGTGGGTGTGTGAGAGGTGTCTGTGAGACACCTTGC	2061	
Db	279	-----CTATAAGTGG	CTGTGTGACACAGCACTCTGTGAGAGTATTTGGAA	CATTCTTGC	330
Oy	2062	CCTCACACAGAGAGCG	GTGAGCACTTCTGTGAGTCCAGCTGCAACATGAGGAAACA	CTGTA	2122
Db	331	TGT---CACACATGGGAC	TGCTCTTCTTGAA	CCCCAGCTGTTCGTGGGAAACATGGGAA	387
Oy	2122	AAAGCAGCAGG	CTCCCGGTGAGGAGCTCTGCTGTGTGCTGCTTCCCATGA-CCACT	2180	
Db	388	GAAAGGAAGAGTGTGTG	-----TGTGTGTGAGTGA	CACTTCCCAAGAGCACT	437
Oy	2181	CCTCTGCTGAATATTA	CTGCTGTAATCTGTGAGACAGATTCGGGTTTATTAAC	CTGCTT	2240
Db	438	CTCTCTTCTGAA	CCCTCACTGATGA-CTGGAGAGGACTCGGGGTTTATTAAC	CTGCTT	496
Oy	2241	TTTATCTGAGAA	CAAGGGTTTGGAAATTAGTGTCTTTTTTCCCACTCCCAAGCTG	2300	
Db	497	TTTATCTGAGAA	CAATGGGTTTGGAACTAGTCTCTTTTCTCCACTTTTACAGAGCTT	556	
Oy	2301	CTCAAGTCA	TTCCACCGGCCCCCTTGAGTGGGACAGGGTAGTGAATCCCGATCCAG	2361	
Db	557	CTCAATTA	TCTCTCCAGGCTTGACTTGGACAGGTAGGGGGGAGACCTGTGGTCCAA	616	
Oy	2361	GAGCTAG	CCCTGACAGAGTGGCTTCCCGTATCCGGTGGGAAAA	CGCCCTGTGCCACAGC	2420
Db	617	GG---TGAC	TGTCCAGGCACTGCCCACTTGTCTAAGAGAAAGGCCCTGTGCTCCAGT	673	
Oy	2421	GGGC	2424		
Db	674	GGAC	677		

RESULT 4

US-08-868-288A-4

Sequence 4, Application US/08868288A

Patent No. 5922567

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Lal, Preci

APPLICANT: Benman, Olga

TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

```

1  SOFTWARE: FastSeq for Windows Version 2.0
2  /
3  / APPLICATION NUMBER: US/08/868.288A
4  /
5  / FILING DATE: June 3, 1997
6  /
7  / PRIOR APPLICATION DATA:
8  / APPLICATION NUMBER:
9  / FILING DATE:
10 / ATTORNEY/AGENT INFORMATION:
11 / NAME: Billings, Lucy J.
12 / REGISTRATION NUMBER: 36,749
13 / REFERENCE/DOCKET NUMBER: PF-0309 US
14 / TELECOMMUNICATION INFORMATION:
15 / TELEPHONE: 415-855-0555
16 / TELEFAX: 415-845-4166
17 / INFORMATION FOR SEQ ID NO: 4:
18 / SEQUENCE CHARACTERISTICS:
19 / LENGTH: 1330 base pairs
20 / TYPE: nucleic acid
21 / STRANDEDNESS: single
22 / TOPOLOGY: linear
23 / IMMEDIATE SOURCE:
24 / LIBRARY: HMT2RAT01
25 / CLONE: 260873
26 /
27 / US-08-868-288A-4
28 /
29 /
30 /
31 /
32 /
33 /
34 /
35 /
36 /
37 /
38 /
39 /
40 /
41 /
42 /
43 /
44 /
45 /
46 /
47 /
48 /
49 /
50 /
51 /
52 /
53 /
54 /
55 /
56 /
57 /
58 /
59 /
60 /
61 /
62 /
63 /
64 /
65 /
66 /
67 /
68 /
69 /
70 /
71 /
72 /
73 /
74 /
75 /
76 /
77 /
78 /
79 /
80 /
81 /
82 /
83 /
84 /
85 /
86 /
87 /
88 /
89 /
90 /
91 /
92 /
93 /
94 /
95 /
96 /
97 /
98 /
99 /
100 /
101 /
102 /
103 /
104 /
105 /
106 /
107 /
108 /
109 /
110 /
111 /
112 /
113 /
114 /
115 /
116 /
117 /
118 /
119 /
120 /
121 /
122 /
123 /
124 /
125 /
126 /
127 /
128 /
129 /
130 /
131 /
132 /
133 /
134 /
135 /
136 /
137 /
138 /
139 /
140 /
141 /
142 /
143 /
144 /
145 /
146 /
147 /
148 /
149 /
150 /
151 /
152 /
153 /
154 /
155 /
156 /
157 /
158 /
159 /
160 /
161 /
162 /
163 /
164 /
165 /
166 /
167 /
168 /
169 /
170 /
171 /
172 /
173 /
174 /
175 /
176 /
177 /
178 /
179 /
180 /
181 /
182 /
183 /
184 /
185 /
186 /
187 /
188 /
189 /
190 /
191 /
192 /
193 /
194 /
195 /
196 /
197 /
198 /
199 /
200 /
201 /
202 /
203 /
204 /
205 /
206 /
207 /
208 /
209 /
210 /
211 /
212 /
213 /
214 /
215 /
216 /
217 /
218 /
219 /
220 /
221 /
222 /
223 /
224 /
225 /
226 /
227 /
228 /
229 /
230 /
231 /
232 /
233 /
234 /
235 /
236 /
237 /
238 /
239 /
240 /
241 /
242 /
243 /
244 /
245 /
246 /
247 /
248 /
249 /
250 /
251 /
252 /
253 /
254 /
255 /
256 /
257 /
258 /
259 /
260 /
261 /
262 /
263 /
264 /
265 /
266 /
267 /
268 /
269 /
270 /
271 /
272 /
273 /
274 /
275 /
276 /
277 /
278 /
279 /
280 /
281 /
282 /
283 /
284 /
285 /
286 /
287 /
288 /
289 /
290 /
291 /
292 /
293 /
294 /
295 /
296 /
297 /
298 /
299 /
300 /
301 /
302 /
303 /
304 /
305 /
306 /
307 /
308 /
309 /
310 /
311 /
312 /
313 /
314 /
315 /
316 /
317 /
318 /
319 /
320 /
321 /
322 /
323 /
324 /
325 /
326 /
327 /
328 /
329 /
330 /
331 /
332 /
333 /
334 /
335 /
336 /
337 /
338 /
339 /
340 /
341 /
342 /
343 /
344 /
345 /
346 /
347 /
348 /
349 /
350 /
351 /
352 /
353 /
354 /
355 /
356 /
357 /
358 /
359 /
360 /
361 /
362 /
363 /
364 /
365 /
366 /
367 /
368 /
369 /
370 /
371 /
372 /
373 /
374 /
375 /
376 /
377 /
378 /
379 /
380 /
381 /
382 /
383 /
384 /
385 /
386 /
387 /
388 /
389 /
390 /
391 /
392 /
393 /
394 /
395 /
396 /
397 /
398 /
399 /
400 /
401 /
402 /
403 /
404 /
405 /
406 /
407 /
408 /
409 /
410 /
411 /
412 /
413 /
414 /
415 /
416 /
417 /
418 /
419 /
420 /
421 /
422 /
423 /
424 /
425 /
426 /
427 /
428 /
429 /
430 /
431 /
432 /
433 /
434 /
435 /
436 /
437 /
438 /
439 /
440 /
441 /
442 /
443 /
444 /
445 /
446 /
447 /
448 /
449 /
450 /
451 /
452 /
453 /
454 /
455 /
456 /
457 /
458 /
459 /
460 /
461 /
462 /
463 /
464 /
465 /
466 /
467 /
468 /
469 /
470 /
471 /
472 /
473 /
474 /
475 /
476 /
477 /
478 /
479 /
480 /
481 /
482 /
483 /
484 /
485 /
486 /
487 /
488 /
489 /
490 /
491 /
492 /
493 /
494 /
495 /
496 /
497 /
498 /
499 /
500 /
501 /
502 /
503 /
504 /
505 /
506 /
507 /
508 /
509 /
510 /
511 /
512 /
513 /
514 /
515 /
516 /
517 /
518 /
519 /
520 /
521 /
522 /
523 /
524 /
525 /
526 /
527 /
528 /
529 /
530 /
531 /
532 /
533 /
534 /
535 /
536 /
537 /
538 /
539 /
540 /
541 /
542 /
543 /
544 /
545 /
546 /
547 /
548 /
549 /
550 /
551 /
552 /
553 /
554 /
555 /
556 /
557 /
558 /
559 /
560 /
561 /
562 /
563 /
564 /
565 /
566 /
567 /
568 /
569 /
570 /
571 /
572 /
573 /
574 /
575 /
576 /
577 /
578 /
579 /
580 /
581 /
582 /
583 /
584 /
585 /
586 /
587 /
588 /
589 /
590 /
591 /
592 /
593 /
594 /
595 /
596 /
597 /
598 /
599 /
600 /
601 /
602 /
603 /
604 /
605 /
606 /
607 /
608 /
609 /
610 /
611 /
612 /
613 /
614 /
615 /
616 /
617 /
618 /
619 /
620 /
621 /
622 /
623 /
624 /
625 /
626 /
627 /
628 /
629 /
630 /
631 /
632 /
633 /
634 /
635 /
636 /
637 /
638 /
639 /
640 /
641 /
642 /
643 /
644 /
645 /
646 /
647 /
648 /
649 /
650 /
651 /
652 /
653 /
654 /
655 /
656 /
657 /
658 /
659 /
660 /
661 /
662 /
663 /
664 /
665 /
666 /
667 /
668 /
669 /
670 /
671 /
672 /
673 /
674 /
675 /
676 /
677 /
678 /
679 /
680 /
681 /
682 /
683 /
684 /
685 /
686 /
687 /
688 /
689 /
690 /
691 /
692 /
693 /
694 /
695 /
696 /
697 /
698 /
699 /
700 /
701 /
702 /
703 /
704 /
705 /
706 /
707 /
708 /
709 /
710 /
711 /
712 /
713 /
714 /
715 /
716 /
717 /
718 /
719 /
720 /
721 /
722 /
723 /
724 /
725 /
726 /
727 /
728 /
729 /
730 /
731 /
732 /
733 /
734 /
735 /
736 /
737 /
738 /
739 /
740 /
741 /
742 /
743 /
744 /
745 /
746 /
747 /
748 /
749 /
750 /
751 /
752 /
753 /
754 /
755 /
756 /
757 /
758 /
759 /
760 /
761 /
762 /
763 /
764 /
765 /
766 /
767 /
768 /
769 /
770 /
771 /
772 /
773 /
774 /
775 /
776 /
777 /
778 /
779 /
780 /
781 /
782 /
783 /
784 /
785 /
786 /
787 /
788 /
789 /
790 /
791 /
792 /
793 /
794 /
795 /
796 /
797 /

```

APPLICATION NUMBER: US/09/235,373  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288  
FILING DATE: June 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-235-373-4

Query Match 1.8%; Score 58; DB 3; Length 1330;  
Best Local Similarity 55.1%; Pred. No. 3.3e-06;  
Matches 134; Conservative 0; Mismatches 103; Indels 6; Gaps 1;

QY 46 ACAATGAAGACTATTCTGCTGTGAGCGGCGAGGAGGCTCTTCTGAAGAGCTGA 105  
DB 155 ACATGGGATTTACTATGAAAGTTCTAGCGCTGAGAGACATGCTCAACCGAGGATATTA 214  
QY 106 AAGCTGCTTACCGGAGGCTCTGTATGCTTACCATCCAGACAGACAGACAGACCCAGAGC 165  
DB 215 AAAGGCAATATCGGAAACTGGCACTGAAGTGCATCCAGATAA-----AAATCTGAGA 268  
QY 166 TCAAGTCACGCGGGAACGACTGTTTAACTTGTTCACAGGCTTATGAAGTCTTAGTG 225  
DB 269 ATAAAGAAAGAGAGAGAGAAATTCAGCAAGTAGCGGAGCATATGAAGTCTGTCCG 328  
QY 226 ACCCCCAACCGAGGCGCATATGATATATATGAGAAAGAGAGAGAGTGAATGGAAGAT 285  
DB 329 ATGCTAAGAAACGGAGCATCTATGACAAATATGCGCAAGAGATTAAATGTGTGGGNG 388  
QY 286 GGG 288  
DB 389 GNG 391

RESULT 6  
US-09-388-993-4  
Sequence 4, Application US/09388993  
Patent No. 6043222  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/388,993  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-388-993-4

Query Match 1.8%; Score 58; DB 3; Length 1330;  
Best Local Similarity 55.1%; Pred. No. 3.3e-06;  
Matches 134; Conservative 0; Mismatches 103; Indels 6; Gaps 1;

QY 46 ACAATGAAGACTATTCTGCTGTGAGCGGCGAGGAGGCTCTTCTGAAGAGCTGA 105  
DB 155 ACATGGGATTTACTATGAAAGTTCTAGCGCTGAGAGACATGCTCAACCGAGGATATTA 214  
QY 106 AAGCTGCTTACCGGAGGCTCTGTATGCTTACCATCCAGACAGACAGACAGACCCAGAGC 165  
DB 215 AAAGGCAATATCGGAAACTGGCACTGAAGTGCATCCAGATAA-----AAATCTGAGA 268  
QY 166 TCAAGTCACGCGGGAACGACTGTTTAACTTGTTCACAGGCTTATGAAGTCTTAGTG 225  
DB 269 ATAAAGAAAGAGAGAGAAATTCAGCAAGTAGCGGAGCATATGAAGTCTGTCCG 328  
QY 226 ACCCCCAACCGAGGCGCATATGATATATATGAGAAAGAGAGAGAGTGAATGGAAGAT 285  
DB 329 ATGCTAAGAAACGGAGCATCTATGACAAATATGCGCAAGAGATTAAATGTGTGGGNG 388  
QY 286 GGG 288  
DB 389 GNG 391

RESULT 7  
US-08-974-546-2  
Sequence 2, Application US/08974546  
Patent No. 5945287  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,546  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:

Query Match	1.8%;	Score 57.2;	DB 2;	length 2349;
Best Local Similarity	56.9%;	Pred. No. 7.8e-06;		
Matches 132;	Conservative	0;	Mismatches 88;	Indels 12;
				Gaps 1;

RESULT 8  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
Datafile: C:\08232463

**ATTORNEY/AGENT INFORMATION:**

Query Match	1.8%;	Score 56.8;	DB 1;	Length 7218;
Best Local Similarity	4.1%;	Pred. No. 2e-05;		
Matches	16;	Conservative	220;	Mismatches 152;
				Indels 0;
				Gaps 0;

Db 1061 AAGTCCCTGACCTGACGCCAAGTCTG 1034

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2



1 CURRENT APPLICATION DATA:  
2 APPLICATION NUMBER: US/09/023,655  
3 FILING DATE: HEREMITH  
4 CLASSIFICATION:  
5 PRIOR APPLICATION DATA:  
6 APPLICATION NUMBER:  
7 FILING DATE:  
8 CLASSIFICATION:  
9 ATTORNEY/AGENT INFORMATION:  
10 NAME: Zeller, Karen J.  
11 REGISTRATION NUMBER: 37,071  
12 REFERENCE/DOCKET NUMBER: PA-0001 US  
13 TELECOMMUNICATION INFORMATION:  
14 TELEPHONE: (650) 855-0555  
15 TELEFAX: (650) 845-4166  
16 INFORMATION FOR SEQ ID NO: 883:  
17 SEQUENCE CHARACTERISTICS:  
18 LENGTH: 1308 base pairs  
19 TYPE: nucleic acid  
20 STRANDEDNESS: single  
21 TOPOLOGY: linear  
22 IMMEDIATE SOURCE:  
23 LIBRARY: GENBANK  
24 CLONE: g1127832  
25  
26 US-09-023-655-883

Query Match	1.5%	Score 48.4	DB 4	Length 1308
Best Local Similarity	52.7%	Pred. NO. 0.0018		
Matches 137, Conservative	0	Mismatches 111	Indels 12	Gaps 1

QY	52	AAAGCTATTTCCTGGTTCGAAACGTGCGAGGAGAGCCCTCTCGAAGAGCTGAAACGTG	111
Db	184	AAGACTTTATTTCGATTTTGGCAATTGGAAAAGAGCTTCAGATGAAATTTTAAAAGG	243
QY	112	CTTACCGGAGGCTCTGATGCTCTTACATCCAGACAGACAGACCAGAGCTCAAGT	171
Db	244	CTTACCGGAAAAACAGGCCCTCAATTTCATCCGGAACAGAAACAAATCTC-----	291
QY	172	CACAGGCGGAACGACTGTTAACTCTGTTCACACAGAGCTTATGAAGTGGTTAGACCCC	231
Db	292	CTCAGCGACAGGAAAAATTTTAAAGAGTGGCAGAAAGCTTTATGAACTATTGATATCTTA	351
QY	232	AAACCAAGGCCATCTATGATATATATGGAAGAGAGGACTGGAAATGGAAGATGGGAGG	291
Db	352	AAAAGAGAGAAATATATGATCAGTTTGGGGAAGAAAGGTTGAAAGAGAGCAAGAGTA	411
QY	292	TTGTGGAAGAGAGGAACCC	311
Db	412	CTGATGACCAAGAGGTACC	431

RESULT 10  
US-09-665-479A-7

```

? Sequence 7, Application US/09665479A
? Patent No. 6673570
? GENERAL INFORMATION:
? APPLICANT: Itoh, Fumiko
? APPLICANT: Itoh, Susumu
? APPLICANT: Heldin, Carl-Henrik
? APPLICANT: ten-Dijke, Peter
? FILE OF INVENTION: SMD ASSOCIATING POLYPEPTIDES
? TITLE REFERENCE: L00461, 70096, US
? CURRENT APPLICATION NUMBER: US/09/665,479A
? CURRENT FILING DATE: 2000-09-20
? PRIOR APPLICATION NUMBER: US 60/154,846
? PRIOR FILING DATE: 1999-09-20
? NUMBER OF SEQ. ID NOS: 20
? SOFTWARE: PatentIn version 3.2
? SEQ. ID NO. 7
? LENGTH: 1318
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: (177)..(1190)
US-09-665-479A-7

```

Query Match	1.5%;	Score 48.4;	DB 4;	Length 1318;
Best Local Similarity	52.7%;	Prod. No. 0.0018;		
Matches 137;	Conservative 0;	Mismatches 111;	Indels 12;	Gaps 12

Qy	52	AAACATTAATACCTGTTGCTGAAACGCGCCAGAGAGGCGCTTCTGGAAGCTGAAAAGCTG	11
Db	184	AAACATATTATTGCACTTTTGGGAATTGAGAAAGAGCTTCAGATGAAGATATTAAAAAG	243
Qy	112	CCTACCGGAGGCTCTGTAATGCTCTACCATCCAGACAGACAGAGACCCAGAGCTCAAT	171
Db	244	CTTACCGGAAAAAAGCGCCCTCAAAATTTCAATCCGGAACAAGAAATATCTC-----	291
Qy	172	CACAGCGGAAAGCATGTTTAACTTGTTCACACAGGCTTATGAAAGTCTTATGATGACCCC	231
Db	292	CTCAGGCAAGGAAAAATTTTAAAGAGGTCGACAAAGCTTATGAAGTATTGATGATCTTA	351
Qy	232	AAACCAAGGCGCATCTATGATATATATATGGAAGAGAGCATGGAAATGGAAGATGGAGG	291
Db	352	AAAAGAGAAATATATATATCATGTTTGGGAGAGAAAGGTGAAAGAGAGCAGAGAGTAA	411
Qy	292	TTTGGAAGAGAGAGAACCC	311
Db	412	CTGATGACAGAGAGGTAC	431

## RESULT 11

```

Sequence 226, Application US/0976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Buchbinder, Jeremy
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 226
LENGTH: 2607
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6673549 138709.5
US-09-976-594-226

```

Query Match	1.5%	Score	48.4	DB	4	Length	2607
Best Local Similarity	52.7%	Pred. No.	0.0027				
Matches	137	Conservative	0	Mismatches	111	Indels	12
						Gaps	1

[illegible]

OY 292 TTGTGAAAGGAGGAAACC 311  
Db 461 CTGATGACAGAGGAGTACC 480

## RESULT 12

US-09-023-655-300  
Sequence 300 Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
NUMBER OF INVENTION: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREMITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 300:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LATRITUT02  
CLONE: 1350210  
US-09-023-655-300

Query Match 1.4%; Score 44.2; DB 4; Length 713;  
Best Local Similarity 61.9%; Pred. No. 0.02; Mismatches 43; Indels 0; Gaps 0;

OY 50 TGAAGACTATTACTGCTGTGTAACGTGCGCAGAGGAGGCTCTTCTGTAAGAGTGAAGC 109  
Db 358 TAAAGATTACTATAGATCTCGGGGGGTAGCAGAGGGGCTCGGATGAGACTGAAGAA 417  
OY 110 TGCTTACCGAGGCTCTGTATGCTCTTACATTCAGACAGACAGACAGACCCAG 162  
Db 418 GGCTTACCGAGAGCTGGCCCTCAATTCCACCAGACCAAGACAGCAGCAGCCTG 470

## RESULT 13

US-09-313-294A-6215  
Sequence 6215 Application US/09313294A  
Patent No. 6476212  
GENERAL INFORMATION:  
APPLICANT: Laligudi, Raghunath V.

APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 6215  
LENGTH: 297  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURES:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6476212 700351424H1  
LOCATION: 4, 10-11, 27, 33-34, 39, 72-88, 114, 263, 274-280  
OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-6215

Query Match 1.3%; Score 42; DB 4; Length 297;  
Best Local Similarity 57.7%; Pred. No. 0.051;  
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 35 GGAGAGCTGGAAGTAAGACTATTACTGCTGTGTAAGTGGCAGAGGAGGCTTCTTC 94  
Db 121 GAAGAGAGCGACACCAAGACTACTGAGATCTCGGGGTGCCAAGTGGGCTCCA 180  
OY 95 TGAAGAGCTGAAGCTGCTTACCGGAGGCTCTGTATGCTTACCATTCAGACAGCAG 154  
Db 181 GAGAGACTCAAGAGGCTTACCGCAGAGGAGGCTATCAAGACACCCGACAGAGGCGG 240  
OY 155 AGACCCAGAG 164  
Db 241 TGAACCCGAG 250

## RESULT 14

US-08-897-340-4  
Sequence 4 Application US/08897340  
Patent No. 595306  
GENERAL INFORMATION:  
APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.  
TITLE OF INVENTION: Weight Control Pathway Genes and Uses  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,340  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/715,032  
FILING DATE: 17-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silveri, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: MNT-005CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-897-340-4

Query Match 1.3%; Score 42; DB 2; Length 1700;  
Best Local Similarity 59.0%; Pred. No. 0.14;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 35 GGAGAGCTGGAACAATGAAGACTATTCTGTTGCTGAACGTGCGAGGAGGAGGCTCTTC 94  
DB 973 GAAGAAGACCAAGAGAAAGATTACTACAAGATCTGGAGTGGACAAAGATGCTCTGA 1032  
QY 95 TGAAGAGCTGAAGAGCTGCTACCGAGGCTCTGTATGCTTACCATCCAGACAAAGACAG 154  
DB 1033 GGACGAGATCAAGAAAGCTTACCGAAACGGGCTTGATGACACCATCCAGATCGGACAG 1092  
QY 155 AG 156  
DB 1093 TG 1094

## RESULT 15

US-09-252-329-4  
Sequence 4, Application US/09252329  
Patent No. 6147192

## GENERAL INFORMATION:

APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.  
TITLE OF INVENTION: Weight Control Pathway Genes and Uses  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHYE & COCKFIELD, LLP

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/252,329

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/897,340

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Silveri, Jean M.

REGISTRATION NUMBER: 39,030

REFERENCE/DOCKET NUMBER: MNI-005CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1700 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-09-252-329-4

QY 35 GGAGAGCTGGAACAATGAAGACTATTCTGTTGCTGAACGTGCGAGGAGGAGGCTCTTC 94  
DB 973 GAAGAAGACCAAGAGAAAGATTACTACAAGATCTGGAGTGGACAAAGATGCTCTGA 1032  
QY 95 TGAAGAGCTGAAGAGCTGCTACCGAGGCTCTGTATGCTTACCATCCAGACAAAGACAG 154  
DB 1033 GGACGAGATCAAGAAAGCTTACCGAAACGGGCTTGATGACACCATCCAGATCGGACAG 1092  
QY 155 AG 156  
DB 1093 TG 1094

Search completed: July 13, 2004, 14:54:04  
Job time : 227 secs

Query Match 1.3%; Score 42; DB 3; Length 1700;  
Best Local Similarity 59.0%; Pred. No. 0.14;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;



QY 181 AACGACTGTTAACTTGTTCACCAAGCTTATGAGTCTTAGTGAACCCCAACCAAGG 240  
| | | | |  
Db 187 AAGGACTGTTAACTTGTTCACCAAGCTTATGAGTCTTAGTGAACCCCAACCAAGG 246  
| | | | |  
QY 241 CCACTATGATATATATGGAAGAGAGACTGGAAATGGAAGATGGAGGTTGTGAAA 300  
| | | | |  
Db 247 CCACTATGATATATATGGAAGAGAGACTGGAAATGGAAGATGGAGGTTGTGAAA 306  
| | | | |  
QY 301 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGTTTGAAGCGGCTGCAGAGAGAGAGAG 360  
| | | | |  
Db 307 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGTTTGAAGCGGCTGCAGAGAGAGAGAG 366  
| | | | |  
QY 361 AGAGAGATTCGACGACGCAACCAATCCAAAGGAAAGATCAGCGTTGAGTAGATGCCA 420  
| | | | |  
Db 367 AGAGAGATTCGACGACGCAACCAATCCAAAGGAAAGATCAGCGTTGAGTAGATGCCA 426  
| | | | |  
QY 421 CCGACCTTTTGTATCGGTATGATGAGAGATGAAAGTGTCCGGCAGTAGTCTTCCGC 480  
| | | | |  
Db 427 CCGACCTTTTGTATCGGTATGATGAGAGATGAAAGTGTCCGGCAGTAGTCTTCCGC 486  
| | | | |  
QY 481 AGATGGAATTTAATTAATGCAATATCCAGTCCATTGAGCACCCTTGACGACGACG 540  
| | | | |  
Db 487 AGATGGAATTTAATTAATGCAATATCCAGTCCATTGAGCACCCTTGACGACGACG 546  
| | | | |  
QY 541 ACAACGCAATCTCTCTGGAAGCCTCTCAAACCAAGATGGAATGGAAGAGTTCCATTA 600  
| | | | |  
Db 547 ACAACGCAATCTCTCTGGAAGCCTCTCAAACCAAGATGGAATGGAAGAGTTCCATTA 606  
| | | | |  
QY 601 ACTTTGGGCTCAGACGAGTAACTTCGGCAAGGGAATGGGAGAGTTGGAAATTTGAGCTG 660  
| | | | |  
Db 607 ACTTTGGGCTCAGACGAGTAACTTCGGCAAGGGAATGGGAGAGTTGGAAATTTGAGCTG 666  
| | | | |  
QY 661 GAGACTTACAGGGGCTTTGTGCGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCT 720  
| | | | |  
Db 667 GAGACTTACAGGGGCTTTGTGCGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCT 726  
| | | | |  
QY 721 TTGTGCAACAACAATGCTCTGCAAGTTTTCATCCGCTGGAATCCGACCGGCTGACCA 780  
| | | | |  
Db 727 TTGTGCAACAACAATGCTCTGCAAGTTTTCATCCGCTGGAATCCGACCGGCTGACCA 786  
| | | | |  
QY 781 CTGTCTTACGTCGGAACTTAGACAAGAACACCGTGGGCTACCTGCAGTGGCCATGGGGTA 840  
| | | | |  
Db 787 CTGTCTTACGTCGGAACTTAGACAAGAACACCGTGGGCTACCTGCAGTGGCCATGGGGTA 846  
| | | | |  
QY 841 TCCAGTACGCAATGAACACTAGCATCGTCCGAGACACTAATAACCAAGCACTTCACTGTG 900  
| | | | |  
Db 847 TCCAGTACGCAATGAACACTAGCATCGTCCGAGACACTAATAACCAAGCACTTCACTGTG 906  
| | | | |  
QY 901 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGCTATCAGACCAAAATTCAG 960  
| | | | |  
Db 907 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGCTATCAGACCAAAATTCAG 966  
| | | | |  
QY 961 ATGACGATCAGACTCGGTGTAAGGATCCCTCAAGAGAGCTTCTTTGGACCGTGTG 1020  
| | | | |  
Db 967 ATGACGATCAGACTCGGTGTAAGGATCCCTCAAGAGAGCTTCTTTGGACCGTGTG 1026  
| | | | |  
QY 1021 AGTACGAGCTGAGAGGAAATCTCCAGGACAGCGTTTGGGTGACGTGTGAGCGTTG 1080  
| | | | |  
Db 1027 AGTACGAGCTGAGAGGAAATCTCCAGGACAGCGTTTGGGTGACGTGTGAGCGTTG 1086  
| | | | |  
QY 1081 GAGTTTCCACGAGGCGTTTCTCTCAAGTCAAGCTCAACAGGCGCAGTCAACATCTCT 1140  
| | | | |  
Db 1087 GAGTTTCCACGAGGCGTTTCTCTCAAGTCAAGCTCAACAGGCGCAGTCAACATCTCT 1146  
| | | | |  
QY 1141 TCCCTATTCACCTGAGGAGACAGCTTCTGCGCAGCGCAATGTTCTATGCAACGTTGGG 1200  
| | | | |  
Db 1147 TCCCTATTCACCTGAGGAGACAGCTTCTGCGCAGCGCAATGTTCTATGCAACGTTGGG 1206  
| | | | |  
QY 1201 CTCTAGTGTCTACTTTGACATGCAACGCTGATCAATCAACCATCACTCAGGGCTCAGA 1260  
| | | | |  
Db 1207 CTCTAGTGTCTACTTTGACATGCAACGCTGATCAATCAACCATCACTCAGGGCTCAGA 1266  
| | | | |  
QY 1261 AAGAGAAAGAAATTTGAGAAAGCAGAGGAAAGCCCGCCACCGATGTGCTGCAGAAAGC 1320  
| | | | |

Db 1267 AAGAGAAAGAAATTTGAGAAAGCAGAGGAAAGCGCGCCACCGATGTGCTCAGAAAGAAC 1326  
| | | | |  
QY 1321 AAGAGCGGAGTCCCGTGTCCGCTGATGACAGAAATCTCTCCGAAGGATTAATTGAGCG 1380  
| | | | |  
Db 1327 AAGAGCGGAGTCCCGTGTCCGCTGATGACAGAAATCTCTCCGAAGGATTAATTGAGCG 1386  
| | | | |  
QY 1381 AAGAGTCCAGAAATGAGGCTCATCATGTCAATGCTGTGATCGGGAAGTTGTCAATGACA 1440  
| | | | |  
Db 1387 AAGAGTCCAGAAATGAGGCTCATCATGTCAATGCTGTGATCGGGAAGTTGTCAATGACA 1446  
| | | | |  
QY 1441 AAGAGCAGAAAGCAGAGAAAGTGAAGTGAATGAAGTGAATGAGTGTCCCTGCAGTGTCTG 1500  
| | | | |  
Db 1447 AAGAGCAGAAAGCAGAGAAAGTGAAGTGAATGAAGTGAATGAGTGTCCCTGCAGTGTCTG 1506  
| | | | |  
QY 1501 TGAAGACCTCGAAAGCTCATCCCTCAAGGAGCCTCCAAAGCTGAGCTGCTGCTTTATG 1560  
| | | | |  
Db 1507 TGAAGACCTCGAAAGCTCATCCCTCAAGGAGCCTCCAAAGCTGAGCTGCTGCTTTATG 1566  
| | | | |  
QY 1561 ACCCGTGTGGGGGGAAGAGAAACCTGAAAGTGTCTATCAGATTCCGGGCGTCTGCTC 1620  
| | | | |  
Db 1567 ACCCGTGTGGGGGGAAGAGAAACCTGAAAGTGTCTATCAGATTCCGGGCGTCTGCTC 1626  
| | | | |  
QY 1621 ATCAGGTGATGTGCTGACAGTGAAGCCCTCCGATACCAAGCAATGCCACAGATCG 1680  
| | | | |  
Db 1627 ATCAGGTGATGTGCTGACAGTGAAGCCCTCCGATACCAAGCAATGCCACAGATCG 1686  
| | | | |  
QY 1681 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGGCAAAAATCTTTCT 1740  
| | | | |  
Db 1687 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGGCAAAAATCTTTCT 1746  
| | | | |  
QY 1741 GGAAGCTCAAAATTTGAAATGAATAAACCCAGACATCAGATGTTTTATTTATTA 1800  
| | | | |  
Db 1747 GGAAGCTCAAAATTTGAAATGAATAAACCCAGACATCAGATGTTTTATTTATTA 1806  
| | | | |  
QY 1801 TTATTTATGAAGGTGATCAATTAATTAATGTAAGGACATGCACACCCAGCTT 1860  
| | | | |  
Db 1807 TTATTTATGAAGGTGATCAATTAATTAATGTAAGGACATGCACACCCAGCTT 1866  
| | | | |  
QY 1861 TTGAAGGTGCTGGGGTGAAGCTGAAGGAGCCCACTGGGAAACCAAGCTGCAGCCCTGGG 1920  
| | | | |  
Db 1867 TTGAAGGTGCTGGGGTGAAGCTGAAGGAGCCCACTGGGAAACCAAGCTGCAGCCCTGGG 1926  
| | | | |  
QY 1921 CATGCTGTTTTCCCAAGGATCAGTTCTGAGAGGAAGGCTCTGAGCCCTGACCTCGCTG 1980  
| | | | |  
Db 1927 CATGCTGTTTTCCCAAGGATCAGTTCTGAGAGGAAGGCTCTGAGCCCTGACCTCGCTG 1986  
| | | | |  
QY 1981 TGTCCGAGCAACGCTGACCGAGCCGCGCTGTAGTTCTTGGCTGGGTGGA 2040  
| | | | |  
Db 1987 TGTCCGAGCAACGCTGACCGAGCCGCGCTGTAGTTCTTGGCTGGGTGGA 2046  
| | | | |  
QY 2041 GGTGCTGAGGAGCAACCTGAGCCCTGACCAAGAGAGTGAAGCACTTGCAGTCCAGC 2100  
| | | | |  
Db 2047 GGTGCTGAGGAGCAACCTGAGCCCTGACCAAGAGAGTGAAGCACTTGCAGTCCAGC 2106  
| | | | |  
QY 2101 TGAACATGAGGAACAACCTGAAGAAGCAGGAGCTCCGCTCAGGAGCCTCTGCTGTG 2160  
| | | | |  
Db 2107 TGAACATGAGGAACAACCTGAAGAAGCAGGAGCTCCGCTCAGGAGCCTCTGCTGTG 2166  
| | | | |  
QY 2161 CTGGCTTCCATGACCACTCTCTCTGCTGAATAATTAATGCTTGAATCTGAGCAGATT 2220  
| | | | |  
Db 2167 CTGGCTTCCATGACCACTCTCTCTGCTGAATAATTAATGCTTGAATCTGAGCAGATT 2226  
| | | | |  
QY 2221 GGGGGTTTATTAACGCTTTTATCTGAGAACAAAGGAGTTTGAATTAATGAGTCTTT 2280  
| | | | |  
Db 2227 GGGGGTTTATTAACGCTTTTATCTGAGAACAAAGGAGTTTGAATTAATGAGTCTTT 2286  
| | | | |  
QY 2281 TTTCCCACTCCCAAGAGCTGCTCAAGTCAATTCACCGGCCCCCTGAGCTTGGGACAGGT 2340  
| | | | |  
Db 2287 TTTCCCACTCCCAAGAGCTGCTCAATTCACCGGCCCCCTGAGCTTGGGACAGGT 2346  
| | | | |  
QY 2341 AGTGTAACTCCGATCCAGAGGCTTAGCCCTGACACAGGTGCTTCCGATCCGAGTGG 2400  
| | | | |

```
Db 2347 AGTGTAACTCCGATCCAGGGGCTAGCCCTGACACAGTGGCTTCCGATCCCGGTG 2406
Qy 2401 GAAATACCCCTGACACAGGGGCTTGAAGCTGGCTGGTCCCTCCACCCGCTGACAC 2460
Db 2407 GAAAAAGCCCTGACACAGGGGCTTGAAGCTGGCTGGTCCCTCCACCCGCTGACAC 2466
Qy 2461 CCACCTCCAGAGTCAAGTCTGGGCAAGGGGAGCTCAAGAGACAGAGACAGGGCTTGG 2520
Db 2467 CCACCTCCAGAGTCAAGTCTGGGCAAGGGGAGCTCAAGAGACAGAGACAGGGCTTGG 2526
Qy 2521 CAAGACATCAGACACACCAACCCAAAGGCTGAGACCCAGGCTCCGCGCTGGTACCA 2580
Db 2527 CAAGACATCAGACACACCAACCCAAAGGCTGAGACCCAGGCTCCGCGCTGGTACCA 2586
Qy 2581 GCAGTGGACATGAGCTCCCGCTCTCTGACAGTCTCAAGCTCTCAAGAGACACAGG 2640
Db 2587 GCAGTGGACATGAGCTCCCGCTCTCTGACAGTCTCAAGCTCTCAAGAGACACAGG 2646
Qy 2641 CCTGTGCTCCGGAGCTTCTTCAAGACCTTCTCAAGCTCCACCTTGGAGTCAAGAT 2700
Db 2647 CCTGTGCTCCGGAGCTTCTTCAAGACCTTCTCTCAAGTCCCACTTGGAGTCAAGAT 2706
Qy 2701 GCAGCGAGCTAGAGACCCCTCAAGGCTGAGCTCTGAGTCAAGTAAAGTTAGTGA 2760
Db 2707 GCAGCGAGCTAGAGACCCCTCAAGGCTGAGCTCTGAGTCAAGTAAAGTTAGTGA 2766
Qy 2761 CCTGTCTCTGGGGGCTTGGAAAGTGGCAAGCTCAAGTCTTGTGTAACCTTGGAGCA 2820
Db 2767 CCTGTCTCTGGGGGCTTGGAAAGTGGCAAGCTCAAGTCTTGTGTAACCTTGGAGCA 2826
Qy 2821 AGCGCCGACAGTGGTGGCTGAGACAGCTGGCGGGGGGGCCCAAGCTGGCGGGCT 2880
Db 2827 AGCGCCGACAGTGGTGGCTGAGACAGCTGGCGGGGGGGCCCAAGCTGGCGGGCT 2886
Qy 2881 CCAGCCACCCCAAGCTGTTGCTGAGTCAAGTCAAGCTCCCAAGCTGATCTGAGT 2940
Db 2887 CCAGCCACCCCAAGCTGTTGCTGAGTCAAGTCAAGCTCCCAAGCTGATCTGAGT 2946
Qy 2941 AAGGCTTAAGAACTCTCTCTCTGTTTGAAGAGCAAGTGGGTTGTCCAAATCTGTA 3000
Db 2947 AAGGCTTAAGAACTCTCTCTCTGTTTGAAGAGCAAGTGGGTTGTCCAAATCTGTA 3006
Qy 3001 ACATTCAATCTCAATTTTAAAGGTTTCTGAGGAGCCCAAGGCGGAGCGGGT 3060
Db 3007 ACATTCAATCTCAATTTT -AAAAAGTTTCTGAGGAGCCCAAGGCGGAGCGGGT 3065
Qy 3061 GAGGCTGTTGATGAGCTGGGCGCCGCGGCTTCCGCTGCGCTGCGCAGTGTCT 3120
Db 3066 GAGGCTGTTGATGAGCTGGGCGCCGCGGCTTCCGCTGCGCTGCGCAGTGTCT 3125
Qy 3121 TCTGGGACCCCATCTCTGCGTTTCATTGTCAGTCTGACAGAGGCACTCAACCA 3180
Db 3126 TCTGGGACCCCATCTCTGCGTTTCATTGTCAGTCTGACAGAGGCACTCAACCA 3185
Qy 3181 ATAAACCTTCTGAAAGCAGAAAAAAA 3213
Db 3186 ATAAACCTTCTGAAAGCAGAAAAAAA 3218
```

```
RESULT 2
US-10-104-047-1796
; Sequence 1796, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIORITY FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 1796
; LENGTH: 2454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1796

Query Match 57.6%; Score 1851.6; DB 16; Length 2454;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1338 GTCGGCTGATGAGGAATGTGTCGAAAGGATTAATGAGCAGAAAGTCCAGAAATGGC 1397
Db 585 GTCGGCTGATGAGGAATGTGTCGAAAGGATTAATGAGCAGAAAGTCCAGAAATGGC 644
Qy 1398 CTATCATGTCTAATGCTTGTACGGGAATTTGTCAATGACAAAGACAGAAAGAGGAG 1457
Db 645 CTATCATGTCTAATGCTTGTACGGGAATTTGTCAATGACAAAGACAGAAAGAGGAG 704
Qy 1458 AAGGTGAAGGTGATGATGACGTCGTCGCTGACAGTGCCTGGTGAAGGACTGGAAGCTC 1517
Db 705 AAGGTGAAGGTGATGATGACGTCGTCGCTGACAGTGCCTGGTGAAGGACTGGAAGCTC 764
Qy 1518 ATCTCAAGAGGCTTCCAAAGGCTGGCTGCTGCTTTTATGACCCGCTGTGGGGAA 1577
Db 765 ATCTCAAGAGGCTTCCAAAGGCTGGCTGCTGCTTTTATGACCCGCTGTGGGGAA 824
Qy 1578 GAGAGAACTGAAAGTGTCTATCAAGTTCGGGGGCTCTGATCAGTGTGTGCTG 1637
Db 825 GAGAGAACTGAAAGTGTCTATCAAGTTCGGGGGCTCTGATCAGTGTGTGCTG 884
Qy 1638 GACAGTGAAGGCTCCGCGATACCAAGCAGTCCCAAGATGATGATGATGATGATGAT 1697
Db 885 GACAGTGAAGGCTCCGCGATACCAAGCAGTCCCAAGATGATGATGATGATGATGAT 944
Qy 1698 GCCAAGAACCAATTTTAAAGGCGCAAAAAATCTTCTCTGAGAGTCAAAATTTG 1757
Db 945 GCCAAGAACCAATTTTAAAGGCGCAAAAAATCTTCTCTGAGAGTCAAAATTTG 1004
Qy 1758 GAAATGAAAAACCCAGACATCAAGTGTATTTATTTATTTATTTATTTATTTATTT 1817
Db 1005 GAAATGAAAAACCCAGACATCAAGTGTATTTATTTATTTATTTATTTATTTATTT 1064
Qy 1818 ACCATTATCAATTAATGAAAGGACATGAGACACCCAGCTTTGAGGGTGCGGGGT 1877
Db 1065 ACCATTATCAATTAATGAAAGGACATGAGACACCCAGCTTTGAGGGTGCGGGGT 1124
Qy 1878 AGGACTAGGCAAGCCCACTGAGAACCAAGACTGAGCTGAGCCATGAGTGTTCCTCA 1937
Db 1125 AGGACTAGGCAAGCCCACTGAGAACCAAGACTGAGCTGAGCCATGAGTGTTCCTCA 1184
Qy 1938 GATCAAGTCTGAGAGGAAGGCTGTGACCTGATCTCGCTGTGTCGAGACACGTG 1997
Db 1185 GATCAAGTCTGAGAGGAAGGCTGTGACCTGATCTCGCTGTGTCGAGACACGTG 1244
Qy 1998 CTGACGCAAGCCCGCGCGCTGTGATGTTCTGGCTGTGAGTGTCTGTGAGAGACC 2057
Db 1245 CTGACGCAAGCCCGCGCGCTGTGATGTTCTGGCTGTGAGTGTCTGTGAGAGACC 1304
Qy 2058 CTGCGCTCAACACAGAGAGCTGAGCACTTCTGACGTCCAGCTGAACATGAGAAACAC 2117
Db 1305 CTGCGCTCAACACAGAGAGCTGAGCACTTCTGACGTCCAGCTGAACATGAGAAACAC 1364
Qy 2178 CTTCTCTCTGTAATTAATGCTGTTGATCTGAGAGAGATGCGGGTTTATTAACCTG 2237
Db 1425 CTTCTCTCTGTAATTAATGCTGTTGATCTGAGAGAGATGCGGGTTTATTAACCTG 1484
Qy 2238 CTTTATCTGAGAAACAAAGGTTTGAATTAATGCTGTTTCTTCCCACTCCAGAG 2297
Db 1485 CTTTATCTGAGAAACAAAGGTTTGAATTAATGCTGTTTCTTCCCACTCCAGAG 1544
```

QY 2298 CTGCTCAAGTCATTCCACGCGGCCCTCGGCTTGAGGACAGGATAGTAACTCCGATCC 2357  
DB 1545 CTGCTCAAGTCATTCCACGCGGCCCTCGGCTTGAGGACAGGATAGTAACTCCGATCC 1604  
QY 2358 CAGGCGCTAGCCTGACACAGGTGGCTTCCGATACCCGATGGGAAAACGCGCTGCAC 2417  
DB 1605 CAGGCGCTAGCCTGACACAGGTGGCTTCCGATACCCGATGGGAAAACGCGCTGCAC 1664  
QY 2418 AGCGGGCTTGAAGCTGGCTGTGTCTCCGACCGGCTTGACACCTCCAGATGACG 2477  
DB 1665 AGCGGGCTTGAAGCTGGCTGTGTCTCCGACCGGCTTGACACCTCCAGATGACG 1724  
QY 2478 TGGCTGGGCAAGGCGACGCTCAAGAGACAGGACCGGCTTGAGGACATCAGACAC 2537  
DB 1725 TGGCTGGGCAAGGCGACGCTCAAGAGACAGGACCGGCTTGAGGACATCAGACAC 1784  
QY 2538 CCAACCCAAAGGCGTGGACCCCAAGGCCCGGCTTGATCCAGAGTGGCACTGCAG 2597  
DB 1785 CCAACCCAAAGGCGTGGACCCCAAGGCCCGGCTTGATCCAGAGTGGCACTGCAG 1844  
QY 2598 TCCCGGCTCTGACAGGTCCAGGCTCTCAAGAGACACGAGGCTGTGCTCCGAGCT 2657  
DB 1845 TCCCGGCTCTGACAGGTCCAGGCTCTCAAGAGACACGAGGCTGTGCTCCGAGCT 1904  
QY 2658 TCCCTGAGACCTTCTCCAGGTGCGCACTTGGGATGACAGATGACGCGGCTAGGAC 2717  
DB 1905 TCCCTGAGACCTTCTCCAGGTGCGCACTTGGGATGACAGATGACGCGGCTAGGAC 1964  
QY 2718 CCCTCCACGCGCTGACCTCGGCTGACGTAAGTTACGTAGGCTGTCTCGGAGGCT 2777  
DB 1965 CCCTCCACGCGCTGACCTCGGCTGACGTAAGTTACGTAGGCTGTCTCGGAGGCT 2024  
QY 2778 GGAAGTGGGACCATGATGCTCTTGCTGACACCTTCGAGGACAGGCGCGCAAGTGG 2837  
DB 2025 GGAAGTGGGACCATGATGCTCTTGCTGACACCTTCGAGGACAGGCGCGCAAGTGG 2084  
QY 2838 GGCCTGAGACGCTGCGCGGCGGAGCCCAAGCTGCGCGGCTTCAGGCCACCAAGCT 2897  
DB 2085 GGCCTGAGACGCTGCGCGGCGGAGCCCAAGCTGCGCGGCTTCAGGCCACCAAGCT 2144  
QY 2898 GTTGTGAAGTCAGGCCAACCTTCCAGCATGTGATCTGATACGCGCTAAGAACCTCC 2957  
DB 2145 GTTGTGAAGTCAGGCCAACCTTCCAGCATGTGATCTGATACGCGCTAAGAACCTCC 2204  
QY 2958 TTCTCTGCTTTGAAAAAGATTTGGGTTGTCCAACTTCTGTAACATTAATCTTCAATTT 3017  
DB 2205 TTCTCTGCTTTGAAAAAGATTTGGGTTGTCCAACTTCTGTAACATTAATCTTCAATTT 2264  
QY 3018 TTAATAAAAGTTTCTGACGCGCCCAAGGCGGAGCGGAGGAGTGTGTTGATG 3077  
DB 2265 TTAATAAAAGTTTCTGACGCGCCCAAGGCGGAGCGGAGGAGTGTGTTGATG 2223  
QY 3078 AGCTTGAGGCGCGGCGCTTCCGATGCGCTTGCAGAGTGTCTTGAGGACCATCTC 3137  
DB 2324 AGCTTGAGGCGCGGCGCTTCCGATGCGCTTGCAGAGTGTCTTGAGGACCATCTC 2283  
QY 3138 TGGCTTTCAATTTGACAGTCTGTAAGAGGACATCAACAAATTAATCTTCTGAAA 3197  
DB 2384 TGGCTTTCAATTTGACAGTCTGTAAGAGGACATCAACAAATTAATCTTCTGAAA 2443  
QY 3198 GCAGAAAAAA 3207  
DB 2444 GCAGAAAAAA 2453

RESULT 3  
US-10-062-674-1802/c  
; Sequence 1802, Application US/10062674  
; Publication No. US20040005559A1  
; GENERAL INFORMATION:  
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.  
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS

FILE REFERENCE: PA-0026-1 CIP  
; CURRENT APPLICATION NUMBER: US/10/062,674  
; CURRENT FILING DATE: 2002-01-30  
; PRIOR APPLICATION NUMBER: US 09/625,102  
; PRIOR FILING DATE: 2000-07-24  
; NUMBER OF SEQ ID NOS: 2217  
; SOFTWARE: PERL Program  
; SEQ ID NO 1802  
; LENGTH: 3600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20040005559A1 253870.6c  
; NAME/KEY: unsure  
; LOCATION: (1) ... (3600)  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-062-674-1802

Query Match 50.1%; Score 1608.8; DB 16; Length 3600;

Best Local Similarity 83.1%; Pred. No. 0; Matches 2839; Conservative 0; Mismatches 363; Indels 213; Gaps 80;

QY 1 GAAAGTTCGAAGATGCGGACGCGCTTGAGCGAGGAGCTGACATGAAGACTATT 60  
DB 3594 GAAAGTTCGAAGATGCGGACGCGCTTGAGCGAGGAGCTGACATGAAGACTATT 3535  
QY 61 ACTCGTGTGAAAGCTGCGGAGGAGGCTC-TTCTGAAGACTGAAGC-TGCCACCG 118  
DB 3534 ACTCGTGTGAAAGCTGCGGAGGAGGCTC-TTCTGAAGACTGAAGC-TGCCACCG 3475  
QY 119 GA--GGCTGTATGCTTAC-ATCCAGACAGACAGAG-CCAGAGCTCAAGTAC 174  
DB 3474 GAAGCTGTATGCTTAC-ATCCAGACAGACAGAG-CCAGAGCTCAAGTAC 3415  
QY 175 AGGCGGACGACTGTTTAACTTGTTCACAG-GCTTATGAAGCTTATGACCCCA 233  
DB 3414 AGGCGGACGACTGTTTAACTTGTTCACAG-GCTTATGAAGCTTATGACCCCA 3355  
QY 234 ACCAGGG-CCATCTATATATATATGAGAGAGAGCTGAAATGAAAGATGGAGGT 292  
DB 3354 ACCAGGGCCATCTATATATATATGAGAGAGAGCTGAAATGAAAGATGGAGGT 3295  
QY 293 TGTGAAAGAGAGAGAACCCCTGC-TGAATTCGAGAGAGTTAGCGGCTG---CAGA 348  
DB 3294 TGTGAAAGAGAGAGAACCCCTGC-TGAATTCGAGAGAGTTAGCGGCTG---CAGA 3235  
QY 349 GAGAGAGAGAGAGAGATTCAGAGCGAACCATCCCAAGGAAAGATCAGCGTTG 408  
DB 3234 GAGAGAGAGAGAGAGATTCAGAGCGAACCATCCCAAGGAAAGATCAGCGTTG 3175  
QY 409 GAGTAGAG-CCACCGACTTTTGAATGCTATGATGAGAGATGAAAGTGTCC-GG 466  
DB 3174 GAGTAGAGTCCACCGACTTTTGAATGCTATGATGAGAGATGAAAGTGTCCGG 3115  
QY 467 CAGTAGCTTT-CCGAGATGAAATTAATAATGACATATCCCAAGTCCATT-GAGGA 524  
DB 3114 CAGTAGCTTTCCGAGATGAAATTAATAATGACATATCCCAAGTCCATTGAGGA 3055  
QY 525 CCCTTGACAGACAGACAGACATCTCTGGAAGGCTTCAACCCAGAAATGA--- 581  
DB 3054 CCCTTGACAGACAGACAGACATCTCTGGAAGGCTTCAACCCAGAAATGA--- 2995  
QY 582 ---AATGAGAGGTTCCATTACTT---CGGCTCAGACGAGTACTTGGCAAGGG 634  
DB 2994 GACAAATGAGAGGTTTCCATTACTTGGCGCTCCAGACGAGTAACTTGGCAAGGG 2935  
QY 635 ATGGGAGAGTTGAAATTTGAGCTGA-----GACCTACAGGGGCTTTTGTGG 685  
DB 2934 ATGGGAGAGCTTGGACATTTGGAGCGCTGAGAACTTACAGGGGCTTTTGTGG 2875  
QY 686 TCTCAAGCTGTCCGTAATCTCAGCAAGATGCTTTGACAAACAACTGTGCTGCA 745



Db	2814	TCCTCAGGTTGTTCCGTAATCTCAACCAAGATGCTTTGTGCAACAAACCTGTGCTCTGCA	2815
Oy	746	GTTT-TCATCCCGTGGAAATCC-GAACCAGCCGTGACCACTGTCTTAAGCTCGGAACCTTAGAC	803
Db	2814	GTTTATCATCCCGTGGAAATCCGTGACCCGGGCTGAGCACTGTCTTAAGCTCGGAACCTTAGAC	2755
Oy	804	AAGAACACCGTGGC-----CTACCTGCAATGGCCGATGGGGATTCAGTCAACCATTAAC	857
Db	2754	AAGAACAACCTGTGGGCTTACCCCTGCAAGTGGGCGATCGGGGATTCACCAATTCCAGTTCCAGCATTAAC	2695
Oy	858	ACTGACATCGTTC-GGAGCACTTAATAACAGGCACTTCACTGTGACCCCTGCAAGCTGGAAAT	916
Db	2694	ACTGACATCGTTCGTAAGGCACTTAATAACAGGCACTTCACTGTGACCCCTGCAAGCTGGAAAT	2635
Oy	917	CCCTCACTCCTTTGCACTGATCAGCTATCAGCAACAATTCCAAAATGACATCAGACTCG	976
Db	2634	CCCTCACTCCTTTGCACTGATCAGCTATCAGCAACAATTCCAAAATGACATCAGACTCG	2575
Oy	977	TGTGAAAGAAATCCCTCAAAAGACGCTTCTTTGGGACGGTGGTGGAGTAACGAGCTGAAG	1036
Db	2574	TGTGAAAGAAATCCCTCAAAAGACGCTTCTTTGGGACGGTGGTGGAGTAACGAGCTGAAG	2515
Oy	1037	GAAATCTCCACGGCAACAGCTTTTGGGTGGACGTGTCAAGCTTGAAGTTCCACAGGCGCT	1096
Db	2514	AGAGATCTCCACGGCAACAGCTTTTGGGTGGACGTGTCAAGCTTGAAGTTCCACAGGCGCT	2455
Oy	1097	TTCTCTCAAAATCAAGCTCAACAGGGCCAGTCAGACATATCTTCTCCATATTCCTTGAC	1156
Db	2454	TTCTCTCAAAATCAAGCTCAACAGGGCCAGTCAGACATATCTTCTCCATATTCCTTGAC	2395
Oy	1157	-GAACAGCCTTC-TGCCAGGCGCAATGTTCTATGACCAAGCGTGGGCGCTTATGATGTCTAC	1214
Db	2394	TGACACAGCCTTCCTGCCCAGGCGCAATGTTCTATGACCAAGCGTGGGCGCTTATGATGTCTAC	2335
Oy	1215	TTTGGCAATGCAACCCGTTC-TGATCATCAAAACCATACCTCAAGGCTCAAGAAAGAAAGAAAT	1273
Db	2334	TTTGGCAATGCAACCCGTCTTATCATCAAAACCATACCTCAAGGCTCAAGAAAGAAAGAAAT	2275
Oy	1274	GGAGAAGCAGAGGGGAAAGCCGCGCAACCGATGTGCTGCAAGAAAGCA-----GAGGC	1327
Db	2274	GGAGAAGCAGAGGGGAAAGCCGCGCAACCGATGTGCTGCAAGAAAGCAAGCAAGAGGC	2215
Oy	1328	GGAATCGGCTGTCCGGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGGCAAGAGATC	1387
Db	2214	GGAATCGGCTGTCCGGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGGCAAGAGATC	2155
Oy	1388	CAGAAATGGGCTCATCATCTGTCAATGCCCTGGTAC-GGGAAATTTGTCAATGACAAGACA	1446
Db	2154	CAGAAATGGGCTCATCATCTGTCAATGCCCTGGTACGGGGAAATTTGTCAATGACAAGAGCA	2095
Oy	1447	GGAGAAGCAGAAAGGTGAAGGTGAATTG--ACGTGACCTGTGCCCTCCAGTCCCTGGTGA	1503
Db	2094	GGAGAAGCAGAAAGGTGAAGGTGAATTGACCTGTGACCTGTGCCCTCCAGTCCCTGGTGA	2035
Oy	1504	AGGACTCGAAGCTCATCTCTCAACGAGG-CTCTCAAGGCTGGGCTGCTGGCTTTATAGA-	1561
Db	2034	AGGACTCGAAGCTCATCTCTCAACGAGGCTCTCTCAAGGCTGGGCTGCTGGCTTTATAGA-	1975
Oy	1562	CCCGTGTGTGGGGGAAAGAAAGAACTGAAATGTCTTATAG-TTCCGGGGCGTCTGTC	1620
Db	1974	CCCGTGTGTGGGGGAAAGAAAGAACTGAAATGTCTTATAGTTCCGGGGCGTCTGTC	1915
Oy	1621	ATCCAGGTGATGG-TGCTGGACAGAGAGGCGCTCCGGATACCAAGCAGT-----CCCA	1672
Db	1914	ATCCAGGTGATGGCTGCTGGACAGAGAGGCGCTTCACTGGATACATATATGACAGTCTCCAT	1855
Oy	1673	CAGGATCGATACAGATGGATAAA--CTGCCAAGAACAGATTTT-AAAAAGCCGCAAAA	1729
Db	1854	CAGGATCGATACAGATGGATAACTTCCCAAGAACAGATTTTCAAAAGCCGCAAAA	1795
Oy	1730	AATC-TTTTCTGGGAGCTTACAAAT-TTGGAAATGAAAAAACCAAGACATCAGATGTT	1786

Db	1734	AATCTGTTTGCATGGGAGCTCAAAATGTTGGAAATGAAAAAACCAGACATCAAGATGTT	1735
OY	1787	TTTATTTTATATATATTTATATATAGAGGTGTAACCATATCA--ATTATGTG--AAGGACAT	1844
Db	1734	TTTATTTTATATATATTTATATATAGAAAGGTGTAACATATACAGATTATGTGAAAGGACAT	1675
OY	1845	GCAGACACCCACAGCTTTTGAAGGTG-CTGGGGGTAGGACTGAGGACAGCCCACTGGGAC	1903
Db	1674	GCAGACACCCCACTTTTAAAGGTGCTGGGGGTAGGACTGAGGCTGCCCTACGGGGTG	1615
OY	1904	CAGACTGACAGCTGCG-----CCATGGCTGTTTCCCAAGATCAG--TTCCTGG	1951
Db	1614	AACCAAGACTGTAGAGCGTGTGTGTCATTTGGCGTGTGTGTGCCAAGGTGTGATGCTTCCGG	1555
OY	1952	AAGGAAGGGCTCTGGCCCC--TGACTCCGCTGTGTCCGAGACAGGTCTGACCCGACG	2008
Db	1554	CGGAAAGGGCTCTGGAACCTGACATCCGGCTGTGTGCCAGACAGCTGTGACCCCAAG	1495
OY	2009	CCGCGCCCTGT--AGTTCTTGGCTGGGTCT--GGAGTGTCTGTGAGACACCTTGCCCT	2064
Db	1494	CCCCGCCCCCTGTAAGTCTTGGCTGGGTCTAGAGAGGTGTCTGTGAGACACCTTGCCCT	1435
OY	2055	CACCAACAGG--AGCGTAGCAATTTGTGC--ATTCACAGCTGAAACAT--GGAAACAACCTG	2120
Db	1434	CACCAACAGAAACGTGAGCACTTCTGCAAGTGTCAAGCTGAACATGAGGAAACAACCTG	1375
OY	2121	AAAAACA--GGCAGGCTCTCCGGTCAGGAG--CTCTGTGTGTGCT--GGCTTCCATGACC	2176
Db	1374	AAABACAATGCAAGGACCTCCGGTCAGGAGACCTCTGTGTGTGTGGCTTCCCATGACC	1315
OY	2177	ACCTCC--TCTGCTGAATATTTACTGTGTTAA--TCTGAGCAGATTGCGGTTTATTA	2232
Db	1314	ACCTCTTCTGTGTAATATTTACTGTGTAATGTGAGCCAGATTCGCGGGTTTATA	1255
OY	2233	AACTGCTTTTATCTGAAACAAG-----GGTTTGA	2267
Db	1254	GAACCTGCTGTGTGAGTCTGGAAGAACACACACCCCGGGGNTTTGAAACAGTTTAG	1195
OY	2268	ATTAGTGTGCTTTTTCCTCCCATCTGCCA--GAGCTGTCAATGTATTCACCG--CCGCC	2323
Db	1194	GTCGGTCTTGTGTTGTCTCCCAAGTCCCAATGAGCTGTCTCAAGTCAATTCACACGACCCCT	1135
OY	2324	TGCGCTTGGGACAGGATGATGTAACTCCG--ATCCAGGGGCTTAGCTGACACAGG	2379
Db	1134	CGGCTTAGGACAGGATGATGTAACTCCGACATCCACAGGGGCTTAGCTGACACAC	1075
OY	2380	TGCGTCTCC-----GTATCCGGGTGGAAAAAGCCCTGTGCCAC--AGCGGCTTGA	2428
Db	1074	TGGGTGGGCTTCCCGGTAACTCCCGTGGGAAAAAGCCCTTGTGCCACACGCGGCTTGA	1015
OY	2429	GCT--GGCGTGTGCTCC--CAACCGGCTGACACCACTTCA--GAGTCAAGTGTGGG	2484
Db	1014	GCTGGGCTGTGTCTCTTACACCGGCTGACACCACTTCAAGGAGTGTCAAGTGTGGG	955
OY	2485	C---AAGGACAGCTTCAAGAG--ACAGGACAGGCGCTTGGCAAGACA---TCAGACA	2534
Db	954	CCAAAGGACAGCTCAAGAGGAGCAGGGGCTTTTGGCAAAAGCATTTACAGACA	895
OY	2535	CACCAACCCAAAGGCTGTGACCCCA--GGCCGGCCCGTGTATCCAGCAGGTGTGACTG	2593
Db	894	CACCAACCCAAAGGCTGTGACCCCAAGGCGCGGCGGTATCCACAGAGGTGTGACTG	835
OY	2594	CAGCTCCCGGCTCTGAGG--TCCAGCGTCTCAAGGAACAACAGGGGCTGTGCTCCG	2652
Db	834	CAGCTCCCGGCTCTGAGGTTTCAAGCGTCTCAAGGAACAACAGGGGCTGTGCTCCG	775
OY	2653	AGCTTCTCTGAGACCTTCTCTCAAGTGTCCCACTTGGAGTGCAGAAATGCAGCGAGCTA	2712
Db	774	AGCTTCTCTGAGACCTTCTCTCAAGTGTCCCACTTGGAGTGTCAAGATGCAGCGAGCTA	715
OY	2713	GG--ACCCCTTCACGAGCTTGACCTGAGCTGACAGTAAATTACGTGAGGCTGTCTCT--	2769
Db	714	GGAGCCCTCTCAACGAGCTGTGAACTCGGCTGTCAATTAAGTTACGTGAGGCTGTCTCTA	655





Dd		1263	AACGGCCCGAGCGCGGAGGTGTGTGTGCATTAAGCTTGGGCCCGGCTTCCTCCGTGC	1322
Oy		3103	GCTCTGCCCGAGAGTGCTTTCTGGGCAACCACCTCCTCGCTTTCATTGGCACTGA	3162
Dd		1323	GCCTTCGCCGAGGTGCTTCTGGGCAACCACCTCCTCGCTTTCATTGGCACTGA	1382
Oy		3163	CAGAAAGGACTCACCAATAAACCCTTCTCGAAAGCAAAAAAAA	3213
Dd		1383	CAGAAGGACTCACCAATAAACCTTCTCGAAAGCAAAAAAAAA	1433
RESULT 5				
US-09-925-300-487				
; Sequence 487, Application US/09925300				
; Patent No. US20020151681A1				
; GENERAL INFORMATION:				
; APPLICANT: Craig Rosen,				
; APPLICANT: Steve Ruben				
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies				
; FILE REFERENCE: PA101				
; CURRENT APPLICATION NUMBER: US/09/925.300				
; CURRENT FILING DATE: 2001-08-10				
; PRIOR APPLICATION NUMBER: PCT/US90/05988				
; PRIOR FILING DATE: 2000-03-08				
; PRIOR APPLICATION NUMBER: 60/124,270				
; PRIOR FILING DATE: 1999-03-12				
; NUMBER OF SEQ ID NOS: 1890				
; SOFTWARE: Patentin Ver. 2.0				
; SEQ ID NO 487				
; LENGTH: 1451				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
; FEATURE:				
; NAME/KEY: misc feature				
; LOCATION: (1256)				
; OTHER INFORMATION: n equals a,t,g, or c				
US-09-925-300-487				
Query Match            44.0%; Score 1413.6; DB 9; Length 1451;				
Best Local Similarity   99.5%; Pred. No. 0;				
Matches 1424; Conservative   3; Mismatches   3; Indels   1; Gaps   1;				
Oy		1783	TGTTTTATTATTATTATTATTATTATATAGAGGTGGTACCATTAATTAATGTGAAGGAC	1842
Dd		1	TGTTTTATTATTATTATTATTATTATAGAGGTGGTACCATTAATTAATGTGAAGGAC	60
Oy		1843	ATGCAGACACCCAGCTTTTAGAGGGTGTGGGGGTAGGACATGAGGACGCCCATCTGGAA	1902
Dd		61	ATGCAGACACCCAGCTTTTAGAGGGTGTGGGGGTAGGACATGAGGACGCCCATCTGGAA	120
Oy		1903	CCAGACTGCAGCTGGCCCCATAGTGCTGTTTTCCCAAGATCAGTTCTTGGAGGAAGGCT	1962
Dd		121	CCAAGCTGCAGCTGGCCCCATAGTGCTTTTTCCCAAGATCAGTTCTTGGAGGAAGGCT	180
Oy		1963	CTGGCCTGACTCCGCTGTGTCCGAGACAAGTGTGACCGGACGCCGCGCTGTAG	2022
Dd		181	CTGGCCTGACTCCGCTGTGTCCGAGACAAGTGTGACCGGACGCCGCGCTGTAG	240
Oy		2023	TTCTTTGGTGGGTCTGGAGGTGTCTGTGGAGACACCTGCCCTTCCACAGAGAGGTGAGC	2082
Dd		241	TTCTTTGGTGGGTCTGGAGGTGTCTGTGGAGACACCTGCCCTTCCACAGAGAGGTGAGC	300
Oy		2083	CACCTTCTGCATTCACCGCTGAACATGGGAAACAACCTGAAAAGAGGACGCTCCGGGT	2142
Dd		301	CACCTTCTGCATTCACCGCTGAACATGGGAAACAACCTGAAAAGAGGACGCTCCGGGT	360
Oy		2143	CAGGAGCCTGTGCTGTGTGCTGAGCTTCCCATGACCACTCTCTGCTGAAATATTACTGC	2202
Dd		361	CAGGAGCCTGTGCTGTGTGCTGAGCTTCCCATGACCACTCTCTGCTGAAATATTACTGC	420
Oy		2203	TTGAATCTGGAGCAGATTGGCGGGTTTATAAACCTGCTTTTATCTGAGAACAAACGGGTT	2262
Dd		421	TTGAATCTGGAGCAGATTGGCGGGTTTATAAACCTGCTTTTATCTGAGAACAAACGGGTT	480

QY	2263	TTGAAATTAGTGGCTTTTTCCTTTTCCCACTCCCGAGAGCTGTCAACTCATTTCCACGGGCCCC	2322
Db	481	TGGAATTATGTCGCTTTTTCCTTCCCACTCCCGAGAGCTGTCAACTCATTTCCACGGGCCCC	540
QY	2323	CTCGGCTTGGGACAGAGGATGTAACTCCGATCCAGGGGCTTAGCCCTGACACAGGTGG	2382
Db	541	CTCGGCTTGGGACAGAGGATGTAACTCCGATCCAGGGGCTTAGCCCTGACACAGGTGG	600
QY	2383	CTTCCCGTATCCCGGTGGAAAAAGCCCTGTGCACACAGGGGCTTAGCTGGCTGTGTCC	2442
Db	601	CTTCCCGTATCCCGGTGGAAAAAGCCCTGTGCACACAGGGGCTTAGCTGGCTGTGTCC	660
QY	2443	CTCCACGGCTGTGACCAACCCACTTCAGAGTGTGAGTGTGGGCAAGGGGACCTCAAGAGG	2502
Db	661	CTCCACGTGTGTGACCAACCCACTTCAGAGTGTGAGTGTGGGCAAGGGGACCTCAAGAGG	720
QY	2503	ACAGAGCCAGGGGCTTGGCAAGATCAGACACCCAAAGGCGTGTGACCCAGG	2562
Db	721	ACAGAGCCAGGGGCTTGGCAAGATCAGACACCCAAAGGCGTGTGACCCAGG	780
QY	2563	CCCGGCTTGTGTACCCAGCAGGTGTGCACTGTCCGCTCTCTGAGGTCCAGGTC	2622
Db	781	CCCGGCTTGTGTACCCAGCAGGTGTGCACTGTCCGCTCTCTGAGGTGTCCAGGTC	840
QY	2623	CTCA CAGGAA CACAGGGGCTGTGCTCCGAGGCTTCCCTTCAGACCCCTCTCCACGGTC	2682
Db	841	CTCA CAGGAA CACAGGGGCTGTGCTCCGAGGCTTCCCTTCAGACCCCTCTCCACGGTC	900
QY	2683	CCACTTGGGATGCAAAATGCAAGCGAGCTAGAGCCCCCTCCACGGCTGTGACCTCGAGCTG	2742
Db	901	CCACTTGGGATGCAAAATGCAAGCGAGCTAGAGCCCCCTCCACGGCTGTGACCTCGAGCTG	960
QY	2743	CAGTAAAGTTACGTGAGGCTGTCTCTCGGGGCTTGGAAATGTGAGCATAGTTGCTCT	2802
Db	961	CAGTAAAGTTACGTGAGGCTGTCTCTCGGGGCTTGGAAATGTGAGCATAGTTGCTCT	1020
QY	2803	TGCTGACCCCTCGAGAGCAAGCGCGCA CAGGTGTGCTGAGACAGCTGGCGGGGGGGC	2862
Db	1021	TGCTGACCCCTCGAGAGCAAGCGCGCA CAGGTGTGCTGAGACAGCTGGCGGGGGGGC	1080
QY	2863	CCCAAGCTCGCGCGGCTCCAGCCCAACCAAGCTGTGCTGAAAGTCAAGGCCAATCTCC	2922
Db	1081	CCCAAGCTCGCGCGGCTCCAGCCCAACCAAGCTGTGCTGAAAGTCAAGGCCAATCTCC	1140
QY	2923	CAGCACTGTGTACTGAGTAAAGGCTAAGAACTCTCTCTGTGTTTGAAGACAGTTC	2982
Db	1141	CAGCACTGTGTACTGAGTAAAGGCTAAGAACTCTCTCTGTGTTTGAAGACAGTTC	1200
QY	2983	GGGTGTGTCAAATTCGTGAACATTCATCCATCTTTTAAAAAAGGTTTCTCTGACGGCCC	3042
Db	1201	GGGTGTGTCAAATTCGTGAACATTCATCTCTCAATTTTTT-AAAAAGTTTTCTCTGACGACC	1259
QY	3043	CACGGCCCGAGCGCGGTGAGCGTGTGTTGCATGAGCTGTGGGCCCGGGCTTCCCGTGC	3102
Db	1260	CACGGCCCGAGCGCGGTGAGCGTGTGTTGCATGAGCTGTGGGCCCGGGCTTCCCGTGC	1319
QY	3103	GCTCTGTGCGGAGGTGCTTCTGGGCAACCATCTCTGTGCTTCAATTGCAGTGCACTGTGA	3162
Db	1320	GCTCTGTGCGGAGGTGCTTCTGGGCAACCATCTCTGTGCTTCAATTGCAGTGCACTGTGA	1379
QY	3163	CAGAAAGGACCTCACACCAATTAACCTTCCGTAAGAGCAGAAAAAAGAAAAA 3213	
Db	1380	CAGAAAGGACCTCACACCAATTAACCTTCCGTAAGAGCAGAAAAAAGAAAAA 1430	



Db 122 GCTTACATCCAGACAGACAGACCCAGAGCTCAAGTCAACAGGCGAAGCACTGTT 181  
Qy 191 TAACCTGTTCACCAAGCTTATGAAGTCTTAAGTGAACCCCAACCAAGGCTCATATGA 250  
Db 182 TAACTTGTTCACCAAGCTTATGAAGTCTTAAGTGAACCCCAACCAAGGCTCATATGA 241  
Qy 251 TATATATGGAG 310  
Db 242 TATATATGGAG 301  
Qy 311 CCCTGCTGAATTCGAG 370  
Db 302 CCCTGCTGAATTCGAG 361  
Qy 371 GCAG 430  
Db 362 GCAG 421  
Qy 431 TGATCGCTATGATGAG 490  
Db 422 TGATCGCTATGATGAG 481  
Qy 491 TAATTAATGAG 550  
Db 482 TAATTAATGAG 541  
Qy 551 CCTCTCTGAG 610  
Db 542 CCTCTCTGAG 601  
Qy 611 CAG 670  
Db 602 CAG 661  
Qy 671 GGGGCTTTGTTCCGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730  
Db 662 GGGGCTTTGTTCCGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721  
Qy 731 AAATCTGTCTCTCAAG 790  
Db 722 AAATCTGTCTCTCAAG 781  
Qy 791 TCGGAGCTTGAAG 850  
Db 782 TCGGAGCTTGAAG 841  
Qy 851 CATGAAG 910  
Db 842 CATGAAG 901  
Qy 911 GGGAGCT 970  
Db 902 GGGAGCT 961  
Qy 971 GACTCGTGTGAAG 1023  
Db 962 GACTCGTGTGAAG 1014

## RESULT 8

US-10-037-270-820  
Sequence 820, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yundong  
APPLICANT: Wang, Dunru  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, John  
APPLICANT: Drmanec, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_fl\_genes Version 1.0  
SEQ ID NO 820  
LENGTH: 1767  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (153)..(1223)  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1767)  
OTHER INFORMATION: n = a, t, c or g  
US-10-037-270-820

Query Match 25.9%; Score 832.2; DB 15; Length 1767;  
Best Local Similarly 99.1%; Pred. No. 3.8e-251;  
Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GAAAGGTGGAG 60  
Db 139 GAAAGGTGGAG 198  
Qy 61 ACTGTTGCTGAAG 120  
Db 199 ACTGTTGCTGAAG 258  
Qy 121 GGTCTGTATGCTTACATCCAG 180  
Db 259 GGTCTGTATGCTTACATCCAG 318  
Qy 181 AAGGAGCTTAACTGTTCAAG 240  
Db 319 AAGGAGCTTAACTGTTCAAG 378  
Qy 241 CCATCTATGATATATATGAG 300  
Db 379 CCATCTATGATATATATGAG 438  
Qy 301 GAG 360  
Db 439 GAG 498  
Qy 361 AG 420  
Db 499 AG 558  
Qy 421 CCAGCTTTTGAATGAG 480  
Db 559 CCAGCTTTTGAATGAG 618  
Qy 481 AGATTGAATTAATTAATGAG 540  
Db 619 AGATTGAATTAATTAATGAG 678  
Qy 541 ACACAGCATCTCTCTGAG 600

Db	679	ACACAGCACCCTCTGTGGAAAGCTTCAACCCAGAAATGGAGAGAGTTCCATTTA	738
Qy	601	ACTTTGGCTCAGACGAGTAACTTGGCGAAAGGATGGGGAGATTGGAATTTGGACCTG	660
Db	739	ACTTTGGCTCAGACGAGTAACTTGGCGAAAGGATGGGGAGAGTTGGAATTTGGACCTG	798
Qy	661	GAGACCTACAGGGGGCCTTTGTCGGGTCGAACGTGTCCGTAATCTCACCAAGATGCT	720
Db	799	GAGACCTACAGGGGGCCTTTGTCGGTCTCAAGCTGTCCGTAATCTCACCAAGATGCT	858
Qy	721	TTGTGACAACAAACTGTGCTGTGCAATTTCAATCCCGTGAATCCGACCCGGCTTGACA	780
Db	859	TTGTGACAACAAACTGTGCTGTGCAATTTCAATCCCGTGAATCCGACCCGGCTTGACA	918
Qy	781	CTGTCTTAAGTCGGAACTTAAGACAGAACACCGTGGGCTAACCTGCAGTGCGAATGGGGTA	840
Db	919	CTGTCTTAAGTCGGAACTTAAGACAGAACACCGTGGGCTAACCTGCAGTGCGAATGGGGTA	978
Qy	841	TTCCAG 845	
Db	979	CCCCG 983	

```

RESULT 9
US-10-117-722-820
Sequence 820, Application US/1011722
Publication No. US20030219744n1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Dmanac, Radje T.
TITLE OF INVENTION: No. US20030219744n1 Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784C1P2BC1P
CURRENT APPLICATION NUMBER: US/10/117,722
CURRENT FILING DATE: 2002-04-04
PRIORITY APPLICATION NUMBER: 09/620,312
PRIORITY FILING DATE: 2000-07-19
PRIORITY APPLICATION NUMBER: 09/552,317
PRIORITY FILING DATE: 2000-04-25
PRIORITY APPLICATION NUMBER: 09/488,725
PRIORITY FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pc_Fl_genes Version 1.0
SEQ ID NO 820
LENGTH: 1767
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (153)..(1223)
NAME/KEY: misc_feature
LOCATION: (1)..(1767)
OTHER INFORMATION: n = a,t,c or g
US-10-117-722-820

```

	Query Match	25.9%	Score 832.3	DB 16	Length 1767
	Best Local Similarity	99.1%	Pred. No. 3.8e-25		
	Matches 837	Conservative 0	Mismatches 8	Indels 0	Gaps 0
QY	1	GAAAGTTGGCAGATGCGCAGCGCCCTTTAGCGAGAGAGCTGGACATGAATCAATT	60		
Db	139	GAAAGTTGGGAAAGATGCGCAGCGCCCTTTAGCGAGAGAGACTGGACATGAATCAATT	198		
QY	61	ACTCGTTGCTGAACGTGGCGAGGAGGCTCTTCTGAAGAGCTGAAGCTGCTTACCGGA	120		
Db	199	ACTCGTTGCTGAACGTGGCGAGGAGGCTCTTCTGAAGAGCTGAAGCTGCTTACCGGA	258		
QY	121	GCGTCTGTATGCTCTTACCATCCAGACAGACAGACCCAGAGCTCAATGTCACAGCGG	180		

Db	259	GGCTCTGTATGCTTACATCCAGACAAGACACAGACCCAGAGCTCAATGCACAGCGG	318
Qy	181	AACGACTGTTTAACTTGTTCACAGGCTTATGAAGTCTTAGTGACCCCAACCAAGG	240
Db	319	AACGACTGTTTAACTTGTTCACAGGCTTATGAAGTCTTAGTGACCCCAACCAAGG	378
Qy	241	CCATCTATATATATATATGGGAAGACAGACTGGAAATGGAAAGATGGGAAGTTGTGGAA	300
Db	379	CCATCTATATATATATATGGGAAGACAGACTGGAAATGGAAAGATGGGAAGTTGTGGAA	438
Qy	301	GGAGGAAACCCCTGCTGAATTCAGAGAGATTGACGGCTTCACAGAGAGAGAAAG	360
Db	439	GGAGGAAACCCCTGCTGAATTCAGAGAGATTGACGGCTTCACAGAGAGAGAAAG	498
Qy	361	AGAGGAGATTGCAGCAGCGAACCAATCCCAAGGGAAGATACGTTGGAGTATGCCA	420
Db	499	AGAGGAGATTGCAGCAGCGAACCAATCCCAAGGGAAGATACGTTGGAGTATGCCA	558
Qy	421	CCGACCTTTTTGATCGCTATGATGAGAGTATGAAGATGTGTCCGGCAGTAGCTTTCCG	480
Db	559	CCGACCTTTTTGATCGCTATGATGAGAGTATGAAGATGTGTCCGGCAGTAGCTTTCCG	618
Qy	481	AGATTGAAATTTAATAAATGACATATATCCAGTCCATTGAGGACCCCTTGACACGACAG	540
Db	619	AGATTGAAATTTAATAAATGACATATATCCAGTCCATTGAGGACCCCTTGACACGACAG	678
Qy	541	ACGACACCATCCTCTCTGGAAGCCTCTACCCCGAAGTGGAAATGGAGGAGTTCCATTA	600
Db	679	ACGACACCATCCTCTCTGGAAGCCTCTACCCCGAAGTGGAAATGGAGGAGTTCCATTA	738
Qy	601	ACTTTGCGCTCAGACGAGTAACTTTCGGCAAGGAGTGGGAGATTGGAAATTTGAGCTG	660
Db	739	ACTTTGCGCTCAGACGAGTAACTTTCGGCAAGGAGTGGGAGATTGGAAATTTGAGCTG	798
Qy	661	GAAACCTAACAGGGGCTTTGTTCGGTCTCAAGCTGTTCCGTATCTCAACCAAGATGCT	720
Db	799	GAAACCTAACAGGGGCTTTGTTCGGTCTCAAGCTGTTCCGTATCTCAACCAAGATGCT	858
Qy	721	TTGTGACCAACAACTGTGCTGCAGTTTCAATCCCGTGGAAATCCGACCGGACTGACCA	780
Db	859	TTGTGACCAACAACTGTGCTGCAGTTTCAATCCCGTGGAAATCCGACCGGACTGACCA	918
Qy	781	CTGTCTTACCTGGAACTTAACAAGAAACCGTGGGCTAACCTGACGTGCGATAGGGGTA	840
Db	919	CTGTCTTACCTGGAACTTAACAAGAAACCGTGGGCTAACCTGACGTGCGATAGGGGCT	978
Qy	841	TTCCAG	845
Db	979	CCCCG	983

```

RESULT 10
US-09-927-738-16
: Sequence 16 Application US/09927738
: Patent No. US20020076799A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongwen
: TITLE OF INVENTION: Compositions and Methods of modulating TGF-B Signaling
: FILE REFERENCE: 17633/1082
: CURRENT APPLICATION NUMBER: US/09/927,738
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: US 60/119786
: PRIOR FILING DATE: 1999-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/03561
: PRIOR FILING DATE: 2000-02-11
: NUMBER OF SEQ ID NOS: 28
: SOFTWARE: Patentin version 3.1
: SEQ ID NO 16
: LENGTH: 844
: TYPE: DNA
: ORGANISM: Unknown
: FEATURE:

```

OTHER INFORMATION: Clone S3 + 125 DNA Sequence  
NAME/KEY: misc feature  
LOCATION: (1)\_(844)  
OTHER INFORMATION: n Can be any nucleotide  
US-09-927-738-16

Query Match 14.2%; Score 456.8; DB 9; Length 844;  
Best Local Similarity 96.9%; Pred. No. 1e-132;

Matches 498; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

```
QY 1350 CAGGAATCTGTCGGAAGATATTGAGCAGAGAAGTCCAGATGCGCTCATCTGTC 1409
DB 1 CAGGAATCTGTCGGA--GATAATTGAGCAGAGAAGTCCAGAAATGGGCTCATCTGTC 59
QY 1410 AATGCTGTGTCGGAAGTGTGTCAATGACAGAGCAGAAAGACGAGAAAGTGAAGTGTG 1469
DB 60 AATGCTGTGTCGGAAGTGTGTGTCAATGACAGAGCAGAAAGACGAGAAAGTGAAGTGTG 119
QY 1470 AATGAGTGAAGTGTGCCCCCTGAGTGTGTCGGAAGACCTGATCTCAGGAG 1529
DB 120 AATGAGTGAAGTGTG--CCCTGAGTGTGCTGGGTAAAGACTGAAGCTCATCTCAG--A 176
QY 1530 GCTTCCAGAGCTGGGCTGCTGCTTTTATGACCCGTGTGTGTGGGGAAGAGAACCTG 1589
DB 177 GCTTCCAGAGCTGGGCTGCTGCTTTTATGACCCGTGTGTGTGGGGAAGAGAACCTG 236
QY 1590 AAATGCTGTGTCGTCCTGGGGGGGCTCTGATCAGGTGATGTGTGTCGACAGTGAAGCC 1649
DB 237 AAATGCTGTGTCGTCCTGGGGGGGCTCTGATCAGGTGATGTGTGTCGACAGTGAAGCC 296
QY 1650 CTCCGATACCAAGACAGTCCACAGATGATACAGATGATTAATCTGCCAAGACGAG 1709
DB 297 CTCCGATACCAAGACAGTCCACAGATGATGATGATTAATCTGCCAAGACGAG 356
QY 1710 ATTTTAAAGGCGCAAAAATCTTTTCTGAGAGTCTACAAATTTGGAATGAAAAA 1769
DB 357 ATTTTAAAGGCGCAAAAATCTTTTCTGAGAGTCTACAAATTTGGAATGAAAAA 416
QY 1770 CCCGACATCAGATGTTTATTTTATTTATTTATAGAAAGTGTGACCATTTATCAAT 1829
DB 417 CCCGACATCAGATGTTTATTTTATTTATTTATAGAAAGTGTGACCATTTATCAAT 476
QY 1830 TATGTGAAGGACATGACAGACCCCGACTTTG 1863
DB 477 TATGTGAAGGACATGACAGACCCCGACTGTG 510
```

## RESULT 11

US-09-884-441-326/c  
Sequence 326, Application US/09884441  
Patent No. US20020119158A1  
GENERAL INFORMATION:  
APPLICANT: Algate, Paul A.  
APPLICANT: Carter, Darick  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.462C7  
CURRENT APPLICATION NUMBER: US/09/884,441  
NUMBER OF SEQ ID NOS: 489  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 326  
LENGTH: 455  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-884-441-326

Query Match 14.2%; Score 455; DB 9; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-132;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1074 AGCGTTGAGTTCACAGGCGGTTTCTCAAGTCAAGCTCAACAGGCGCATGACACA 1133
DB 1133 TACTTCTTCCCTATTTCACAGGCGGTTTCTCAAGTCAAGCTCAACAGGCGCATGACACA 1133
```

```
DB 455 AGCGTTGAGTTCACAGGCGGTTTCTCAAGTCAAGCTCAACAGGCGCATGACACA 396
QY 1134 TACTTCTTCCCTATTTCACAGGCGGTTTCTCAAGTCAAGCTCAACAGGCGCATGACACC 1193
DB 395 TACTTCTTCCCTATTTCACAGGCGGTTTCTCAAGTCAAGCTCAACAGGCGCATGACACC 336
```

```
QY 1194 GTGGGCGCTTAGTGTCTATTGTCATGACCGGTGTGATCAATCAACCATGCTCAGG 1253
DB 335 GTGGGCGCTTAGTGTCTATTGTCATGACCGGTGTGATCAATCAACCATGCTCAGG 276
```

```
QY 1254 GCTCAGAAAGAGAGATTTGAGAGACAGAGGAAAGCCGCCACCGATGTGCTCAG 1313
DB 275 GCTCAGAAAGAGAGATTTGAGAGACAGAGGAAAGCCGCCACCGATGTGCTCAG 216
```

```
QY 1314 AAGAGCAAGAGCGGAGTCCGCTGTCGCTGATGACAGAAATCTGTCCAGAGATATT 1373
DB 215 AAGAGCAAGAGCGGAGTCCGCTGTCGCTGATGACAGAAATCTGTCCAGAGATATT 156
```

```
QY 1374 GAGGCAAGAGTCCAGAAATGGGCTCATCATGTCAATGCCGTACGGGAAGTTTGTG 1433
DB 155 GAGGCAAGAGTCCAGAAATGGGCTCATCATGTCTCATGTGTAAGGAAATTTGTG 96
```

```
QY 1434 AATGACAAAGACAGAGAGACGAGAGAGTGAAGTGAATGACGTGACCTGTGCCCTGACAG 1493
DB 95 AATGACAAAGACAGAGAGACGAGAGAGTGAAGTGAATGACGTGACCTGTGCCCTGACAG 36
```

```
QY 1494 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```

```
QY 1528 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1528
DB 35 TGCTGTGAGAGACTGGAAGTCTCATCTCAGCA 1
```



Db 335 GTGGGGCTCTAGTGGTCTACTTCCATCCACCGTCTGATCTCAACCATCTCAGG 276  
Qy 1254 GCTCAGAAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 113  
Db 275 GCTCAGAAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 216  
Qy 1314 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 1373  
Db 215 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 156  
Qy 1374 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 1433  
Db 155 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 96  
Qy 1434 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 1493  
Db 95 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 36  
Qy 1494 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1528  
Db 35 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1

## RESULT 13

US-09-827-271-326/c  
; Sequence 326, Application US/09827271  
; Publication No. US20030165504A1  
; GENERAL INFORMATION:  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.462C6  
; CURRENT APPLICATION NUMBER: US/09/827,271  
; NUMBER OF SEQ ID NOS: 461  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 326  
; LENGTH: 455  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-827-271-326

## Query Match

14.2%; Score 455; DB 10; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-132;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1074 AGCGTTGAGTTCACAGGCGCTTCTCTCAAAAGTCAAGTCAACAGGCGCAGTCAAGCA 1133  
Db 455 AGCGTTGAGTTCACAGGCGCTTCTCTCAAAAGTCAAGTCAACAGGCGCAGTCAAGCA 396  
Qy 1134 TACTTCTTCCCTATTTCACCTTGAAGGACAGCTTCTGCGGCGCAATGTTCTATGCGACC 1193  
Db 395 TACTTCTTCCCTATTTCACCTTGAAGGACAGCTTCTGCGGCGCAATGTTCTATGCGACC 336  
Qy 1194 GTGGGCGCTCTAGTGTCTACTTGGCCATGACCGTCTGATCATCAAAACCATACCTCAG 1253  
Db 335 GTGGGCGCTCTAGTGTCTACTTGGCCATGACCGTCTGATCATCAAAACCATACCTCAG 276  
Qy 1254 GCTCAGAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 1313  
Db 275 GCTCAGAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 216  
Qy 1314 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 1373  
Db 215 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 156  
Qy 1374 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 1433  
Db 155 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 96  
Qy 1434 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 1493

Db 95 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 36  
Qy 1494 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1528  
Db 35 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1

## RESULT 14

US-10-198-053-326/c  
; Sequence 326, Application US/10198053  
; Publication No. US20030124140A1  
; GENERAL INFORMATION:  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Hill, Paul  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.462C9  
; CURRENT APPLICATION NUMBER: US/10/198,053  
; NUMBER OF SEQ ID NOS: 624  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 326  
; LENGTH: 455  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-053-326

## Query Match

14.2%; Score 455; DB 15; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-132;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1074 AGCGTTGAGTTCACAGGCGCTTCTCTCAAAAGTCAAGTCAACAGGCGCAGTCAAGCA 1133  
Db 455 AGCGTTGAGTTCACAGGCGCTTCTCTCAAAAGTCAAGTCAACAGGCGCAGTCAAGCA 396  
Qy 1134 TACTTCTTCCCTATTTCACCTTGAAGGACAGCTTCTGCGGCGCAATGTTCTATGCGACC 1193  
Db 395 TACTTCTTCCCTATTTCACCTTGAAGGACAGCTTCTGCGGCGCAATGTTCTATGCGACC 336  
Qy 1194 GTGGGCGCTCTAGTGTCTACTTGGCCATGACCGTCTGATCATCAAAACCATACCTCAG 1253  
Db 335 GTGGGCGCTCTAGTGTCTACTTGGCCATGACCGTCTGATCATCAAAACCATACCTCAG 276  
Qy 1254 GCTCAGAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 1313  
Db 275 GCTCAGAAAGAGAAATTTGAGAGACAGAGGAAAGCCCGCACCCGATGCTGCAG 216  
Qy 1314 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 1373  
Db 215 AAGAACGACAGAGGCGAGTCCGCTGTCGGCTGATGACAGAAATCTGTCCGAGATTAAT 156  
Qy 1374 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 1433  
Db 155 GAGGCAAGAAAGTCCAGAAATGGGCTCATATGCTCAATGCTGTGACGGAAGTTGTC 96  
Qy 1434 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 1493  
Db 95 AATGACAAAGACGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTGGCCCTGCAG 36  
Qy 1494 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1528  
Db 35 TGCCTGTGAAGACTGGAAGCTCATCTCAGCGA 1

## RESULT 15

US-09-764-868-167  
; Sequence 167, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 167
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (454)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (465)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-167
```

Query Match 14.1%; Score 452; DB 9; Length 479;

Best Local Similarity 98.1%; Pred. No. 2.3e-131;

Matches 471; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

```
Qy 18 GCGACGGCTTGAGCGAGAGAGCTGGACAATGAACTATTACTGTTGCTGAAGCTG 77
Db 1 GCGACGGCTTGAGCGAGAGAGCTGGACAATGAACTATTACTGTTGCTGAAGCTG 60
Qy 78 GCGAGGAGGCGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGAGGCTCTGTATGCTTAC 137
Db 61 GCGAGGAGGCGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGAGGCTCTGTATGCTTAC 120
Qy 138 CATCCAGACAGACACAGAGACCCAGAGCTCAAGTCAAGCGGAAACGACTGTTAACTT 197
Db 121 CATCCAGACAGACACAGAGACCCAGAGCTCAAGTCAAGCGGAAACGACTGTTAACTT 180
Qy 198 GTTACCGAGGCTTATGAAGTGTAGTGACCCCAACGAGGGCATGTATGATATATAT 257
Db 181 GTTACCGAGGCTTATGAAGTGTAGTGACCCCAACGAGGGCATGTATGATATATAT 240
Qy 258 GGGAGAGAGAGACTGGAATGGAAGATGGAGGTTGTGAAGAGAGAGAACCCCTGCT 317
Db 241 GGGAGAGAGAGACTGGAATGGAAGATGGAGGTTGTGAAGAGAGAGAACCCCTGCT 300
Qy 318 GAAATTGAGAGAGAGTTGAGCGGCTGCAAGAGAGAGAGAGAGAGAGATTGCAGCAG 377
Db 301 GAAATTGAGAGAGAGTTGAGCGGCTGCAAGAGAGAGAGAGAGAGAGATTGCAGCAG 360
Qy 378 CGAACCAATCCAGAGGAAAGATCAGCGTGGAGTAGATGCCAGACCTTTTGTATGCC 437
Db 361 CGAACCAATCCAGAGGAAAGATCAGCGTGGAGTAGATGCCAGACCTTTTGTATGCC 420
Qy 438 TATGATGAGAGATGAAAGATGTCCCGCAGTAGCTTT-CCGCAATTGAAATTAAATA 496
Db 421 TATGATGAGAGATGAAAGATGTCCCGCAGTAGCTTTCCGCAATTGAAATTAAATA 479
```

Search completed: July 13, 2004, 18:19:52

Job time : 1509 secs



**This page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:46:53 ; Search time 7858 Seconds  
(without alignments)  
12210.128 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213  
Sequence: 1 gaaagctgcgaagatgcgcg.....gaaagcagaaaaaaaaa 3213

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc1:\*  
9: gb\_estc1:\*  
10: gb\_estc2:\*  
11: gb\_htc:\*  
12: gb\_estc3:\*  
13: gb\_estc4:\*  
14: gb\_estc5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vtc:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1471.8	45.8	2926	11 AK053156	AK053156 Mus muscu
2	953.4	29.7	1033	12 BM909566	BM909566 AGENCOURT
3	927.4	28.9	1019	12 BM927425	BM927425 AGENCOURT
4	921	28.7	1005	9 AL552317	AL552317 AL552317

5	917	28.5	1027	13 BQ066706	BQ066706 AGENCOURT
6	914.8	28.5	972	9 AL525212	AL525212 AL525212
7	882.6	27.5	972	9 AL562459	AL562459 AL562459
8	876.4	27.3	1019	13 BQ061784	BQ061784 AGENCOURT
9	869	27.0	1025	12 BQ054445	BQ054445 AGENCOURT
10	863	26.9	1053	13 BQ063535	BQ063535 AGENCOURT
11	861.6	26.8	930	13 BQ058032	BQ058032 AGENCOURT
12	858	26.7	935	13 BQ050592	BQ050592 AGENCOURT
13	857.2	26.7	1017	13 BQ062383	BQ062383 AGENCOURT
14	856.6	26.7	1018	13 BQ063825	BQ063825 AGENCOURT
15	852.8	26.5	989	9 AL575646	AL575646 AL575646
16	852.2	26.5	1124	13 BQ061457	BQ061457 AGENCOURT
17	851.6	26.5	3844	11 AK089983	AK089983 Mus muscu
18	850.2	26.5	955	13 BQ0712401	BQ0712401 AGENCOURT
19	847	26.4	924	13 BQ0390960	BQ0390960 AGENCOURT
20	840.4	26.2	1201	9 AL582666	AL582666 AL582666
21	836.4	26.0	896	12 B1906525	B1906525 603064011
22	833.4	25.9	1012	13 BQ061825	BQ061825 AGENCOURT
23	833.2	25.9	869	12 BM468897	BM468897 AGENCOURT
24	827.8	25.8	997	13 BQ056258	BQ056258 AGENCOURT
25	825.8	25.7	1201	9 AL561468	AL561468 AL561468
26	822.4	25.6	990	12 BM927596	BM927596 AGENCOURT
27	821.6	25.6	962	13 BQ057649	BQ057649 AGENCOURT
28	820.4	25.5	987	13 BQ056732	BQ056732 AGENCOURT
29	816.6	25.4	823	12 B1256757	B1256757 602977620
30	816	25.4	891	12 BM013882	BM013882 603639484
31	815.4	25.4	969	13 BQ056772	BQ056772 AGENCOURT
32	812.4	25.3	825	9 AL040438	AL040438 DKF2p434C
33	810.6	25.2	969	13 BX350638	BX350638 BX350638
34	810	25.2	887	13 BQ187518	BQ187518 AGENCOURT
35	807.4	25.1	904	13 BQ056088	BQ056088 AGENCOURT
36	806.6	25.1	934	13 BQ056939	BQ056939 AGENCOURT
37	803.8	25.0	949	13 BQ0711401	BQ0711401 AGENCOURT
38	800.8	24.9	920	13 BQ0646154	BQ0646154 AGENCOURT
39	800.6	24.9	1031	13 BQ062757	BQ062757 AGENCOURT
40	800	24.9	938	13 BQ0643192	BQ0643192 AGENCOURT
41	796.8	24.8	1201	13 BX332157	BX332157 BX332157
42	796.4	24.8	964	13 BQ0642728	BQ0642728 AGENCOURT
43	792.8	24.7	1057	13 BQ063938	BQ063938 AGENCOURT
44	791.4	24.6	904	14 CA489230	CA489230 AGENCOURT
45	791.2	24.6	1001	13 BQ066712	BQ066712 AGENCOURT

## ALIGNMENTS

RESULT 1  
AK053156 2926 bp mRNA linear HTC 20-SEP-2003  
DEFINITION Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched  
library, clone:R030019A03 product:hypothetical DnaJ N-terminal  
domain containing protein, full insert sequence.

ACCESSION AK053156  
VERSION AK053156.1 GI:26343260  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PubMed 10349636

REFERENCE  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PubMed 11042159

REFERENCE  
AUTHORS  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Komori, H., Akiyama, J., Nishii, K., Kikunishi, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
20530913  
11076861

REFERENCE  
AUTHORS  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL  
AUTHORS  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
AUTHORS  
6  
(bases 1 to 2926)  
Adachi, S., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, K., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku, K.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Yamamoto, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers

FEATURES  
source  
1. .2926  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:E030019A03"  
/db\_xref="MGI:2424521"  
/db\_xref="Taxon:10090"  
/clone="E030019A03"  
/tissue\_type="lung"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
16. .1695  
/note="unnamed protein product; hypothetical DnaU  
N-terminal domain containing protein (InterPro|IPR001623,  
evidence: InterPro)  
putative"  
/codon\_start=1  
/protein\_id="BAC35287.1"  
/db\_xref="MGI:26343261"

/translation="MATAISEEELNEDYVSLNVRREASSEELKAAYRLCMLYHPD  
KRPDEPLKQAKRFLVHQAYEVSDPQTRAIYIDYKRGKEMEMEVKERTPAE  
IREEPRILOREBERREIIOORTNPKGTISGVNATLPDYREDEVEDVSSGSPQIEIN  
KHHSIOSIEAIPITADITALLSSISTONONGGGSYNFALRPTASKMELEIFGAGDL  
QSLPGLKLRNLTRECFVTYALQIOPSSRGIRPGITVYLANLNDKNTGVYQMRGTI  
QAMMTSIVTRDKTCHFYVALQIGPHSFLSIVHKFDDQTRVSKLKGFPETI  
VYGGAEKTSRHSVIGAAVSIQVPGVSKVLNLSASQYFPPIHLITQDLPSPAFA  
TGPVIVYAVYARLIRPYLRAQKEKELEKRENTASDIQKQEAFAVRLEQSEVR  
RIIEEESMGILIVNMYGKRVNKRNEKRVKVIDVPLQICVKSGLITLTSKSK  
ALRPGYDPCVCEEKSRLVLYPQVRLHVMYVDSBALRIRPQSRIDTDG"

polyA\_signal  
2900..2905  
/note="putative"  
2926  
polyA\_site  
/note="putative"

ORIGIN  
Query Match 45.8%; Score 1471.8; DB 11; Length 2926;  
Best Local Similarity 79.0%; Pred. No. 0;  
Matches 1917; Conservative 0; Mismatches 437; Indels 73; Gaps 11;

QY 1 GAAAGGTGCGAAGATGCGAGCGGCTTGAGGAGGAGGAGGCTGACATGAAGACTATT 60  
DB 2 GAAAGGTGCGAAGATGCGAGCGGCTTGAGGAGGAGGAGGCTGACATGAAGACTATT 61  
QY 61 ACTCGTGTGCTGAACGTCGCGCAGGAGGCTCTTCTTGAAAGGCTGAAGCTGCTACCGGA 120  
DB 62 ACTCACTGTGTAACGTCGCGCAGGAGGCTCTTCTTGAAAGGCTGAAGCTGCTACCGGA 121  
QY 121 GGTCTGTATGCTTACATCCAGACACACAGACCCAGAGCTCAAGTCAACAGCGG 180  
DB 122 GGTGTGTATGCTTACATCCAGACACAGACCCAGAGCTCAAGTCAACAGCGG 181  
QY 181 AACGATGTTTAACTGTTTCAACGAGCTTATGAAGTGTGTGACCCCAACAGG 240  
DB 182 AGCGGTGTTTAACTGTTTCAACGAGCTTATGAAGTGTGTGACCCCAACAGG 241  
QY 241 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAAGTTGGAAA 300  
DB 242 CTATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAAGTTGGAAA 301  
QY 301 GGAAGGAAACCCCTGCTGAATTCAGAGAGAGTTGAGCGCTGCAGAGAGAGAGAG 360  
DB 302 GAAAGGAAACCCCTGCTGAATTCAGAGAGAGTTGAGCGCTGCAGAGAGAGAGAG 361  
QY 361 AGAGAGATGTCAG 420  
DB 362 AGAGAGATGTCAG 421  
QY 421 CCGACCTTTTATGCTGATGATGAGAGATGAGAGATGATGATGATGATGATGATGAT 480  
DB 422 CTGACCTTTTATGCTGATGATGAGAGATGAGAGATGATGATGATGATGATGATGAT 481  
QY 481 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGGACCTTTGACAGAGAG 540  
DB 482 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGGACCTTTGACAGAGAG 541  
QY 541 AACAGGCAATCTCTCTGGAAGCTCTCAACCAAGATGGAATGGAAGAGAGTCCATTA 600  
DB 542 AACAGGCAATCTCTCTGGAAGCTCTCAACCAAGATGGAATGGAAGAGAGAGTCCATTA 601  
QY 601 ACTTTGCGCTCAGACAGATTAATTCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 602 ACTTTGCGCTCAGACAGATTAATTCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661  
QY 661 GAGACCTCAAGGAG 720  
DB 662 GAGACCTCAAGGAG 721  
QY 721 TTGTGCAACAACAGTGTCTGAGAGTTTCAATCCGTGGAATTCGACCCGAGCTGACCA 780  
DB 722 TTGTGCAACAACAGTGTCTGAGAGTTTCAATCCGTGGAATTCGAGCTGAGCTGACCA 781  
QY 781 CTGTCTAGCTGGAACCTAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840

D	782	CTGTGCTAGCTCGGAACTTGGACAAAGAACTGTGGCTACTGTGAGTGGGCTG	841
Q	841	TTCCATGACGCTAGTAACAATGATCGTCCGAGACATTAACAGGACCTTCACTGTGG	900
D	842	TTCCATGACGCTAGTAACAATGATCGTCCGAGACATTAACAGGACCTTCACTGTGG	901
Q	901	CCCTGACGCTGGGAAATCCCTCACTCTTGGACATGATCAGATACCAAAATTTCCAG	960
D	902	CTTTCAGCTGGGATCCCTCACTCTTGGACATGATCAGATACCAAAATTTCCAG	961
Q	961	ATGACATCAGACTCTGTGAAAGATCCCTCAAGGAGCTTCTTGGAGCGTGTGG	1020
D	962	ACGACGACGAGCTCGAGTAAAGGATCCCTCAAGGAGCTTCTTGGAGCACTATGG	1021
Q	1021	AGTACGAGCTGAGAGAAATCTCCAGGACACGCTTTTGGTGCAGCTGTGCGTTG	1080
D	1022	AGTATGAGCTGAGAGAAATTTCCGACACAGCTTGGGTGGCTGTGAGCATTTG	1081
Q	1081	GAGTTCCACAGGGGCTTCTCTCAAGTCAAGCTCAAGGCGCAGTCAATCTTCT	1140
D	1082	GCGTCCCTCAGGGGTGTCTCTCAAGTCAAGCTCAAGGCGCAGTCAATCTTCT	1141
Q	1141	TTCCCTATTCATTTGACGAGACGCTTCCGACGCGCATGTCTTATGCCACGTTGGGCG	1200
D	1142	TTCCCTATTCATTTGACGAGTACGCTGCTCCGACGCGCTGTGTCTTACGCCACGTTGGGCG	1201
Q	1201	CTCTAGTGTCTACTTTTCCATGACGCTGTGATCATCAACATACCTTCAGGCTCAGA	1260
D	1202	CGCTGTGTCTACTTTTCCATGACGCTGTGATCATCAACATACCTTCAGGCTCAGA	1261
Q	1261	AAGAGAGAAATTTGAG	1320
D	1262	AAGAGAGAAATTTGAG	1321
Q	1321	AAG	1380
D	1322	AAG	1381
Q	1381	AAG	1440
D	1382	AAG	1441
Q	1441	AGAGCAG	1500
D	1442	AGAGCAG	1501
Q	1501	TGAAGAGCTGAGAGCTCATCTCTCAGGAGGCTCCAGAGGCTGGGCTGCTTTATG	1560
D	1502	TGAAGAGCTCAGAGCTCATCTCTCAGAGGCTCCAGAGGCTGGGCTGCTTTATG	1561
Q	1561	ACCGGTGTGGGGAG	1620
D	1562	ATCATATGTGTGGGGAG	1621
Q	1621	ATCAGGTGATGT	1680
D	1622	ACCGGTGTGTGGGGAG	1681
Q	1681	ATACAGATGATTAATGCTG-CAGAGACAGATTTTAAAGGCGCAAAAATCTTTCC	1739
D	1682	ACACAGACGGGTAAACCGCTTGGGAACTCGAGAGAAAAGAGGCTAGAGAAACCTTTTC	1741
Q	1740	TGGAGTCTCAAAATTTGGAA-ATGAAAAAACCCAGACATCAGATGTTTTATTTAT	1798
D	1742	TAAAGAGCTCAAAATTTGGAAAGCAGGAAAAACCCAGACATGAGATGTTTTATTTAT	1801
Q	1799	TATATATATAG	1858
D	1802	CTCCA-----GAAAGGGGGAGCTGTATCAATTAATTAAGAGACATCAGACCTTAC	1855
Q	1859	TTTTAGGGGTGCTGGGGGTAGAGACTGAGGAGCCCACTGGGAAACAGACTGAGCTGG	1918

D	1856	TTCCATGATGCTGT-GGGGTAG	1914
Q	1919	CCCATGAGCTGTTTTTCCAGAGATCATGTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1978
D	1915	TC-----TTCGCCAAGAGATCATGTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1959
Q	1979	TGTGTCCGAGACACAGTGTGACCGAG	2038
D	1960	-----CTCATAGT	1988
Q	2039	GAGGTGTGTGTGAGACACCTGTGCTTCAACAG	2098
D	1989	TGAGAGTATTTGAAATCATTTGTGCTCA---CAGATGGAGCTGCTTCTTCAAGAGCCAT	2045
Q	2099	GCTGAACATGGGAAACATCTGAAAAG	2158
D	2046	GCTGTGTGTGTGAGAAACATGGGAAAG	2095
Q	2159	TGCTGTGTGTGTGAG-CCAGCTCTCTGTGCTGAAATATTAATCTTGAATCTGAGAGAG	2217
D	2096	AGTACATTTCCAG	2154
Q	2218	ATTGCGGTTTATTAATGCTGTTTTATCTGAG	2277
D	2155	ACTGCGGTTTATTAATGCTGTTTTATCTGAG	2214
Q	2278	TTTTTTTCCCATCTCCAG	2337
D	2215	TTCTTCCCATTTTATCAAG	2274
Q	2338	GGTATGTTATCTCCAG	2397
D	2275	GGGGGGAG	2331
Q	2398	TGGGAAAAAG	2424
D	2332	AG	2358

RESULT 2	BM909566	1033 bp	mRNA	linear	EST 12-MAR-2002
LOCUS	BM909566				
DEFINITION	AGENCOURT_6640947 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5434173				
ACCESSION	BM909566				
VERSION	BM909566.1	GI:19359945			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1033)				
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@pds-remail.nih.gov Tissue Procurement: Lou Staudt CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LNCM910 row: m column: 22 High quality sequence stop: 733. Location/Qualifiers 1..1033 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5434173" /tissue_type="lymphoma, cell line"				



Db	121	GTATCTCTTACCATCCAGACCAAGACCCAGAGCTCAAGTCAAGGGGAAACAC	180
Oy	187	TGTTTAACCTTGTTCACCAAGGCTTATGAAATGCTTAACTCCCAACCAAGGCATCT	246
Db	181	TGTTTAACCTTGTTCACCAAGGCTTATGAAATGCTTAACTCCCAACCAAGGCATCT	240
Oy	247	ATGATATATATATGGGAAGAGAGACTGGAATATGGAAGATGGAGGTTTGGAAAGAGAG	306
Db	241	ATGATATATATATGGGAAGAGAGACTGGAATATGGAAGATGGAGGTTTGGAAAGAGAG	300
Oy	307	GAACCCCTGCTGAATATGGAGAGAGTTTGAAGCGGCTGAGAGAGAGAGAGAGAGAG	366
Db	301	GAACCCCTGCTGAATATGGAGAGAGTTTGAAGCGGCTGAGAGAGAGAGAGAGAGAG	360
Oy	367	GATTGCACAGCGCAACCAATCCCAAGGAAACGATCAGCGTTGAGATGACACCGAC	426
Db	361	GATTGCACAGCGCAACCAATCCCAAGGAAACGATCAGCGTTGAGATGACACCGAC	420
Oy	427	TTTTTGAATCGCTATATGAGAGAGTATGAAGATGTGTCCGGCAGTATGCTTTCCGAGATTG	486
Db	421	TTTTTGAATCGCTATATGAGAGAGTATGAAGATGTGTCCGGCAGTATGCTTTCCGAGATTG	480
Oy	487	AAATTAATTAATAATGACATATCCGAGTCCATTGAGAGCAACCTTGCACACGACAG	546
Db	481	AAATTAATTAATAATGACATATCCAGTCCATTGAGAGCAACCTTGCACACGACAG	540
Oy	547	CCATCCTCTCTGGAGAGCCTCTCAACCCAGAAATGAGAAATGAGAGAGTTTCAATTAATTG	606
Db	541	CCATCCTCTCTGGAGAGCCTCTCAACCCAGAAATGAGAAATGAGAGAGTTTCAATTAATTG	600
Oy	607	CGCTCAGACGAGTAACTTCCGCAATGGAGTGGGAGAGTGGAAATTTGAGAGCTGGAGCC	666
Db	601	CGCTCAGACGAGTAACTTCCGCAATGGAGTGGGAGAGTGGAAATTTGAGAGCTGGAGAC	660
Oy	667	TACAGGGGCGCTTGTGGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGGA	726
Db	661	TACAGGGGCGCTTGTGTGGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGGA	720
Oy	727	CAACAAATGTGCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGGGCTGACACTGTCC	786
Db	721	CAACAAATGTGCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGGGCTGACACTGTCC	780
Oy	787	TAGCTCGGAACCTTAGACAAGAACACCGTGGGCTACTGCACT-GGCGATGGGGATATCCAG	845
Db	781	TAGCTCGGAACCTTAGACAAGAACACCGTGGGCTACTGCAAGTGGGGATATCCAG	840
Oy	846	TCAGCGATGAACACTAGACATGTGTCCGAGACACTAAACCAAGCACTTCACTGTGGCCCTG	905
Db	841	TCAGCGATGAACACTAGACATGTGTCCGAGACACTAAACCAAGCACTTCACTGTGGCCCTG	900
Oy	906	CAGCTGGGAATCCCTCAGTCCCTTT-GCACTGATCGCTATCAGC-ACAAATTTCCAAGATG	963
Db	901	CAGCTGGGAATCCCTCAGTCCCTTTGGACCGATACGCTATACGACAAATTTCCAAGATG	960
Oy	964	ACGA-----TCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGAGCGGTG	1016
Db	961	ACGAATCCGACTTGTGTGGAAAGGATCCCTCAAGCAGGCTTCTTTGTGGGGG	1019
RESULT 4			
AL552317			
LOCUS	1005 bp	mRNA	linear
DEFINITION	AL552317 Homo sapiens PLACENTA COT 25-NORMALIZED	Homo sapiens	EST 31-MAY-2003
ACCESSION	AL552317		
VERSION	AL552317.2		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 1005)		

AUTHORS L. W. B., Gruber C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12891103.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 271.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS01D1069CH11Q1&cluster=271.r. Contact :  
 Feng Liang Email : fliang@liletech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS01D1069CH11Q1.  
 location/Qualifiers  
 1..1005  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS01D1069YP21"  
 /issue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Query Match 28.7%; Score 921; DB 9; Length 1005;  
 Best Local Similarity 99.5%; Pred. No. 1.2e-228;  
 Matches 932; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```

OY      19  CGACGGCCCTTAGGAGGAGGAGCGTCGACATATGAAGACTATTACTCGTGTGTAACGCG 78
DB      68  CGACGGCCTTAGGAGGAGGAGCGTCGACATATGAAGACTATTACTCGTGTGTAACGCG 127
OY      79  GCAGGAGAGCGCTCTTCTGTAAGAGCTGTAAGCTGCTACCGGAGCCTGTATGCTTACC 138
DB      128  GCAGGAGAGCGCTCTTCTGTAAGAGCTGTAAGAGCTGCTACCGGAGCCTGTATGCTTACC 187
OY      139  ATCCAGACACAGACAGAGACCAGAGCTCAAGTCAAGTCACAGGCGGAAAGACTGTTTAACTTG 198
DB      188  ATCCAGACACAGACAGAGACCAGAGCTCAAGTCAAGTCACAGGCGGAAAGACTGTTTAACTTG 247
OY      199  TTCAACAGAGCTTATGAAGTGTCTATGTATGATACCCCAACCAAGGCCATCTATATATATATG 258
DB      248  TTCAACAGAGCTTATGAAGTGTCTATGTATGATACCCCAACCAAGGCCATCTATATATATATG 307
OY      259  GGAAGAGAGAGACTGGAATATGAAAGATGCGAGGTTGTGAAAGAGAGGAAAGCCCTGCTG 318
DB      308  GGAAGAGAGAGACTGGAATATGAAAGATGCGAGGTTGTGAAAGAGAGGAAAGCCCTGCTG 367
OY      319  AAATTGAGAGAGATTGAGCGAGCTGCAGAGAGAGAGAAAGAGAGAGATTTGACGAGC 378
DB      368  AAATTGAGAGAGATTGAGCGAGCTGCAGAGAGAGAGAAAGAGAGATTTGACGAGC 427
OY      379  GAACCAATCCCAAGAGAAACATCAAGCGTTGAGATGATGCCACCGACTTTTGTATGCT 438
DB      428  GAACCAATCCCAAGAGAAACATCAAGCGTTGAGATGATGCCACCGACTTTTGTATGCT 487
OY      439  ATGATGAGAGATATGAAGATGTGTCGCGGAGTACTCTTCCGCAATTTGAATTAATATAA 498
DB      488  ATGATGAGAGATATGAAGATGTGTCGCGGAGTACTCTTCCGCAATTTGAATTAATATAA 547
OY      499  TGCAATATCCCACTGCAATTGAGGACACCTTTGACAGGACAGACACAGCCATCCTCTCTG 558
DB      548  TGCAATATCCCACTGCAATTGAGGACACCTTTGACAGGACAGACACAGCCATCCTCTCTG 607
OY      559  GAAGCTTCAACCCAGATATGAATATGAGAGAGGTTCCATTAACTTTGCGTCAAGCAG 618
DB      608  GAAGCTTCAACCCAGATATGAATATGAGAGAGGTTCCATTAACTTTGCGTCAAGCAG 667
  
```

Qy	619	TAACATCGGCAAAAGGAGATGGGGAGAGTTGGAAATTTGGAGCTGGAGAACCTACACAGGGGACCTT	678
Db	668	TAACCTTCGGCAAAAGGAGATGGGGAGAGTTGGAAATTTGGAGCTGGAGAACCTACACAGGGGACCTT	727
Qy	679	TGTTTGGGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA CAACAACCTGTG	738
Db	728	TGTTTGGGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA CAACAACCTGTG	787
Qy	739	CTGTCGAGTTTTTCAATCCCGTGGAAATCCGACCCGGGCTGACCACTGTCTAGCTGGGAAC	798
Db	788	CTCTGCAATTTTTCAATCCCGTGGAAATCCGACCCGGGCTGACCACTGTCTAGCTGGGAAC	847
Qy	799	TAGACAAAGAACCCGTGGCTTACTGTCAAGTGGCGATGGGCTATTCACAGTCAAGCATGAACA	858
Db	848	TAGACAAAGAACCCGTGGCTTACTGTCAAGTGGCGATGGGCTATTCACAGTCAAGCATGAACA	907
Qy	859	CTAGCATGTGCCAGACACTAAACCAACCACTTCACTGTGGCCCTGAGCT -GGGAATC	917
Db	908	CTAGCATGTGCCAGACACTAAACCAACCACTTCACTGTGGGCCCTGAGCTGGGGAAATC	967
Qy	918	CCTCACTCCTTTTGCACTGATCAAGCTATCAGCACAAT	954
Db	968	CCTCACTCCTTTTGCACTGATCAAGCTATCAGCACAAT	1004

RESULT 5	
BQ066706	
LOCUS	
DEFINITION	BQ066706 1027 bp mRNA linear EST 02-APR-2007
ACCESSION	AGNCOCRT_6869906 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5931375 5' mRNA sequence.
VERSION	BQ066706 BQ066706
	GI:19895752

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE 1 (pages 1 to 1027)  
AUTHORS NIH-MGC <http://mgc.nhl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: L12M212 row: j column: 16  
High quality sequence stop: 676.

## FEATURES

### Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5931375"
/tissue="lymphoma, cell line"
/lab_host="RDH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/name="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

```

Query Match	28.5%	Score 917;	DB 13;	Length 1027;
Best Local Similarity	96.1%;	Pred. No. 1.4e-227;		
Matches 983; Conservative	0;	Mismatches 35;	Indels 5;	Gaps 4;

OY	TTGGGAAGATGGCGACCGGCTTGTGAGCGAGAGAGGCTGTGACAAATGAAGAATAATTAACTGCT	66
Db	1 TTGGGAAGATGGCGACCGGCTTGTGAGCGAGAGAGGCTGTGACAAATGAAGAATAATTAACTGCT	60
OY	67 TGCTGAACGTGCGCAGGAGGAGGCTTCTTCTGAAGAAGCTGAAGAGCTGCTCAACCGAGGCTCT	126
Db	61 TGCTGAACGTGCGCAGGAGGAGGCTTCTTCTGAAGAAGCTGAAGAGCTGCTCAACCGAGGCTCT	120
OY	127 GTATGCTCTACATTCAGACAAAGACAAAGACCCAGAGCTCAATGTCAAGGCGGAACAGAC	186
Db	121 GTATGCTCTACATTCAGACAAAGACAAAGACCCAGAGCTCAATGTCAAGGCGGAACAGAC	180
OY	187 TGTTTAAACCTGTGTTCACCAAGGCTTATGAAGTGCTTATGATGCCCAACACAGGCGCATCT	246
Db	181 TGTTTAAACCTGTGTTCACCAAGGCTTATGAAGTGCTTATGATGCCCAACACAGGCGCATCT	240
OY	247 ATGATATATATGGGAAGAGAGACTGGAATAATGGAAGATGGAGAGTTGTGTGAATGAAGAGA	306
Db	241 ATGATATATATGGGAAGAGAGACTGGAATAATGGAAGATGGAGAGTTGTGTGAATGAAGAGA	300
OY	307 GAACCCCTGCTGGAATTCAGAGAGAGTTTGAAGGCGCTGCAGAGAGAGAGAGAAAGAGAGA	366
Db	301 GAACCCCTGCTGGAATTCAGAGAGAGTTTGAAGGCGCTGCAGAGAGAGAGAGAAAGAGAGA	360
OY	367 GATTGACAGACAGACCAATCCCAAGGAGAAAGATCAGCGTTGTGAGTAGATGCAACCGACC	426
Db	361 GATTGACAGACAGACCAATCCCAAGGAGAAAGATCAGCGTTGTGAGTAGATGCAACCGACC	420
OY	427 TTTTGTGATCGGTATGATGAAGAGATATGAAGATGTGTCCGCGCATATGCTTTCCGAGATTG	486
Db	421 TTTTGTGATCGGTATGATGAAGAGATATGAAGATGTGTCCGCGCATATGCTTTCCGAGATTG	480
OY	487 AAATTAATTAATAATGCAATATCCCAATGTCATTTGAGGACACCTTGACAGCGACAGACACAG	546
Db	481 AAATTAATTAATAATGCAATATCCCAATGTCATTTGAGGACACCTTGACAGCGACAGACACAG	540
OY	547 CCATCTCTCTGGAAGGCTCTGCAACCCAGAAATGGAATGAGAGGTTCCATTAACTTTG	606
Db	541 CCATCTCTCTGGAAGGCTCTGCAACCCAGAAATGGAATGAGAGGTTCCATTAACTTTG	600
OY	607 CGCTCAGACGATTAATCTTCGCGAAAGGAGATGGGAGAGTTGGAATTTGAGAGCTGGAAGACC	666
Db	601 CGCTCAGACGATTAATCTTCGCGAAAGGAGATGGGAGAGTTGGAATTTGAGAGCTGGAAGACC	660
OY	667 TACAGGAGGCTTGTGCTCGATCTCAAGCTGTGTTCCGTAATCTCAACAACAAGATGCTTTGTGA	726
Db	661 TACAGGAGGCTTGTGCTCGATCTCAAGCTGTGTTCCGTAATCTCAACAACAAGATGCTTTGTGA	720
OY	727 CAACAAACTGTGCTCTGTGAGTTTCAATCCGCTGGAATCCGACCCGCGCTGACCATCTGTCC	786
Db	721 CAACAAACTGTGCTCTGTGAGTTTCAATCCGCTGGAATCCGACCCGCGCTGACCATCTGTCC	780
OY	787 TAGCTTCGGAACCTTGAACAAGAACACCGTGGGCTTCTGCAAGT--GGGATGGGGTATATCA	844
Db	781 TAGCTTCGGAACCTTGAACAAGAACACCGTGGGCTTCTGCAAGTGGCGAATGGGGTATATCA	840
OY	845 GTGAGCCATG-AACTAGCATGTGTCCGAGACAT-AAAAACAGCCCATTTCACTGTGGCC	902
Db	841 GTGAGCCATGAAACATAGCATGTGTCCGAGACATTAATAAAACAGGCCCATTTACTGTGGCC	900
OY	903 CTGAGAGCTGGAAATCCCTCACTCC--TTTGAATGTGATAGATATGAGACAAATTCAGAGA	961
Db	901 CTGAGAGCTGGAAATCCCTCACTCCCTTTTGTCCCGATATGATATCAGACCAATTCAGAGA	960
OY	962 TGAGGATTCAGACTGTGTGAAAGAGATCCCTTAAAGCAGGCTTCTTTTGGGACGCGTGTGGA	1021
Db	961 TGAGGATTCAGACTGTGTGAAAGAGATCCCTTAAAGCAGGCTTCTTTTGGGACGCGGCGT	1020
OY	1022 GTA 1024	



Db 1021 GGA 1023

RESULT 6  
AL525212  
LOCUS  
DEFINITION AL525212 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
CDNA clone CS0DC005YN12 5-PRIME, mRNA sequence.

ACCESSION AL525212  
VERSION AL525212.2  
KEYWORDS GI:31043467  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
L1,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 13, 2001 this sequence version replaced gi:12788705.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 271.r For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC005D06NP1&cluster=271.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC005D06NP1.  
Location/Qualifiers  
1..972  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC005YN12"  
/issue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_1ib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 28.5%; Score 914.8; DB 9; Length 972;  
Best Local Similarity 99.8%; Pred. No. 5e-227;  
Matches 916; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 11 GAAGATGGCGAGCGCTTTGAGCGAGAGAGCTGAGCAATGAACTATTACTGCTGT 70  
55 GGATATGGCGAGCGCTTTGAGCGAGAGAGCTGAGCAATGAACTATTACTGCTGT 114  
71 GAAGTGGCGAGAGCGCTTTGAGAGAGCTGAAAGCTCTTACCGGAGGCTCTGTAT 130  
115 GAAGTGGCGAGAGCGCTTTGAGAGAGCTGAAAGCTCTTACCGGAGGCTCTGTAT 174  
131 GCTTACCATCCAGACAGACAGAGAGAGAGCTCAAGTCAAGCGAGCGAGAGAGCTT 190  
175 GCTTACCATCCAGACAGAGAGAGAGAGAGCTCAAGTCAAGCGAGCGAGAGCTT 234  
191 TAACTTGTTCACGAGGCTTATGAAGTGCTTAGTGAACCCCAAAACAAGGCCATCTATGA 250  
235 TAACTTGTTCACGAGGCTTATGAAGTGCTTAGTGAACCCCAAAACAAGGCCATCTATGA 294  
251 TATATATGGAG 310  
295 TATATATGGAG 354  
311 CCCTGTGAATTTGATT 370

Db 355 CCCTGTGAATTTGATT 414  
371 GCAGCAG 430  
415 GCAGCAG 474  
431 TGATCGCTATGATGATT 490  
475 TGATCGCTATGATGATT 534  
491 TAAATGAATGACATATCCAGTTCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550  
535 TAAATGAATGACATATCCAGTTCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 594  
551 CCTTCTGGAAGCTCTCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 610  
595 CCTTCTGGAAGCTCTCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654  
611 CAGAGAGTAAGTTCGAG 670  
655 CAGAGAGTAAGTTCGAG 714  
671 GAGGCTTTGTCGAGTCAAGCTGTCCGTAAATCTCACACCAAGATGCTTTGAGACAAC 730  
715 GAGGCTTTGTCGAGTCAAGCTGTCCGTAAATCTCACACCAAGATGCTTTGAGACAAC 774  
731 AAATGCTCTGAGATTTTCAATCCGTTGAATCCAGCCGCGCTGACACTGTCTTAC 790  
775 AAATGCTCTGAGATTTTCAATCCGTTGAATCCAGCCGCGCTGACACTGTCTTAC 834  
791 TCGGAACCTGAG 850  
835 TCGGAACCTGAG 894  
851 CATGAACATGAGATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 910  
895 CATGAACATGAGATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954  
911 GGGATTCCTCACTCCTT 928  
955 GGGATTCCTCACTCCTT 972

Db 911 GGGATTCCTCACTCCTT 928  
955 GGGATTCCTCACTCCTT 972

RESULT 7  
AL562459/c  
LOCUS  
DEFINITION AL562459 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
CDNA clone CS0DC005YN12 3-PRIME, mRNA sequence.

ACCESSION AL562459  
VERSION AL562459.2  
KEYWORDS GI:31286471  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
L1,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 15, 2001 this sequence version replaced gi:12910899.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 271.r For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC005D06NP1&cluster=271.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC005D06NP1.  
Location/Qualifiers



source 1. .972  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC005YN12"  
/issue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_id="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 27.5%; Score 882.6; DB 9; Length 972;  
Best Local Similarity 93.4%; Pred. No. 1.3e-218;  
Matches 898; Conservative 37; Mismatches 24; Indels 2; Gaps 2;  
QY 1251 AGGCTCAGAAAGAAAGAAATTGGAGACAGAGGAAAGCCCGCCAGCTGTCG 1310  
DB 972 AGGCTCAGAAAGAAAGAAATT-GAAGAAARARRAARAAAMCMAMMAAARAAAA 914  
QY 1311 CAGAAAGACAGAGGCGAGTCCGCTGTCGCGCTGATGAGGAATCTGTCGAGGAT 1370  
DB 913 MAAGAAAGAAAGAGGCGAGTCCAMMAMACGCTGAGRAATCTGTCGAGAGAT 854  
QY 1371 ATTGAGCAGAAAGTCCAGATAGGCTCATCATGTCATGCTGTCGAGGAGTTT 1430  
DB 853 ATTGAGCAGAAAGTCCAGATAGGCTCATCATGTCATGTCATGTCGAGGAGTTT 794  
QY 1431 GTCAATGACAAAGCAGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTCCCTG 1490  
DB 793 GTCAATGACAAAGCAGAGAAAGACGAGAAAGTGAAGTGAATGACGTAAGTCCCTG 734  
QY 1491 CAGTGCCTGAGTGAAGACTGGAAGCT-CATCTCAGAGAGGCTCCAGAGGCTGGGCTGC 1549  
DB 733 CAGTGCCTGAGTGAAGACTGGAAGCTGGAAGCTGGAAGGCTCCAGAGGCTGGGCTGC 674  
QY 1550 TGGCTTTATGACCCGCTGTGGGAGAAAGAAAGAAAGTGTCTATCATGTTCCG 1609  
DB 673 TAGATTTATTAACCCGCTGTGGGAGAAARARAAAGAAAGTGTCTATCATGTTCCG 614  
QY 1610 GGGCTCTGCAATCAGGTATGCTGTGACAGTGAAGCCCTCCGATACCAAGACATC 1669  
DB 613 GGGCTCTGCAATCAGGTATGCTGTGACAGTGAAGCCCTCCGATACCAAGACATC 554  
QY 1670 CCACAGGATGATACAGATGATTAATGCAAGAAACAGATTTTAAAGGCGGCAAAA 1729  
DB 553 CCACAGGATGATACAGATGATTAATGCAAGAAACAGATTTTAAAGGCGGCAAAA 494  
QY 1730 AATCTTTCTGGAAGCTACAAATTTGAAATGAAAAACCCAGACATCAGATGTTT 1789  
DB 493 AATCTTTCTGGAAGCTACAAATTTGAAATGAAAAACCCAGACATCAGATGTTT 434  
QY 1790 AATTTATATTAATTAATTAAGAGTGTACATTAATTAATTAATTAATTAATTAATTA 1849  
DB 433 AATTTATATTAATTAATTAAGAGTGTACATTAATTAATTAATTAATTAATTAATTA 374  
QY 1850 CACCCAGCTTTGAGGGTCTGAGGGTAGAGCTGAGGAGCCCACTGGGAACAGACT 1909  
DB 373 CACCCAGCTTTGAGGGTCTGAGGGTAGAGCTGAGGAGCCCACTGGGAACAGACT 314  
QY 1910 GAGGCTGAGCCATGAGCTGTTTCCAGAGATCAGTTCTGAGGAGGAGGCTGAGCC 1969  
DB 313 GAGGCTGAGCCATGAGCTGTTTCCAGAGATCAGTTCTGAGGAGGAGGCTGAGCC 254  
QY 1970 TGACTCCGCTGTGTCCGAGACACAGTGTGACCGGAGCCGCTGTATTTCTG 2029  
DB 253 TGACTCCGCTGTGTCCGAGACACAGTGTGACCGGAGCCGCTGTATTTCTG 194  
QY 2030 CTGGGCTGAGAGTGTGTGAGACACCTGAGCTCAGCAAGAGAGTGAAGCACTTCT 2089  
DB 193 CTGGGCTGAGAGTGTGTGAGACACCTGAGCTCAGCAAGAGAGTGAAGCACTTCT 134

QY 2090 GAGTCCAGCTGAACATGGAGAAACACCTGAAAGACAGGAGGCTCCGCTGAGGAG 2149  
DB 133 GAGTCCAGCTGAACATGGAGAAACACCTGAAAGACAGGAGGCTCCGCTGAGGAG 74  
QY 2150 CCTGTCTGT 2209  
DB 73 CCTGT 14  
QY 2210 T 2210  
DB 13 T 13

RESULT 8  
B0061784  
LOCUS B0061784 1019 bp mRNA linear EST 02-APR-2002  
DEFINITION AGENCOURT\_6854890 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5921279  
5', mRNA sequence.  
ACCESSION B0061784  
VERSION B0061784.1 GI:19886171  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1019)  
NIH-MGC http://mhc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNC2086 row: e column: 24  
High quality sequence strop: 596.  
FEATURES  
Location/Qualifiers  
1..1019  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5921279"  
/issue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

ORIGIN

Query Match 27.3%; Score 876.4; DB 13; Length 1019;  
Best Local Similarity 96.8%; Pred. No. 5.4e-217;  
Matches 937; Conservative 0; Mismatches 26; Indels 5; Gaps 4;  
QY 7 TTGCGAAGATGGCGAGCGCTTGAACGAGAGAGCTGGAACAATGAAGATTTACTGT 66  
DB 1 TTGCGAAGATGGCGAGCGCTTGAACGAGAGAGCTGGAACAATGAAGATTTACTGT 60  
QY 67 TGTGAAAGCTGAGAGGAGGCTCTTCTGAAGAGCTGAAGCTGTAACGAGAGCTCT 126  
DB 61 TGTGAAAGCTGAGAGGAGGCTCTTCTGAAGAGCTGAAGCTGTAACGAGAGCTCT 120  
QY 127 GTATGCTTAACATCCAGCAAGACAGAGACCCAGAGCTCAAGTCAAGGCGGAAGAC 186





[illegible]

Qy	1213	TTTAAAGGCGCGAAAAAATCTTTTCTGGGAGCTACAAATTTGGAAATGAAAAAACC	1772
Db	360	TTTAAAGGCCCGCAAAAATCTTTTCTGGGAGCTACAAATTTGGAAATGAAAAAACC	419
Qy	1773	AGACATCAGATGTTTTATTTATTTATATATTTATATAGAAAGTGGTACATTTATCAATTTAT	1832
Db	420	AGACATCAGATGTTTTATTTATTTATTTATTTATTTATTAATAAGGTGTACATTTATCAATTTAT	479
Qy	1833	GTGAAAGGACATGCACACACCCCAAGCTTTTGAAGGCTCTGGGGGTAGAGACTGAGCGACCC	1892
Db	480	GTGAAAGGACATGCACACACCCCAAGCTTTTGAAGGCTCTGGGGGTAGAGACTGAGCGACCC	539
Qy	1893	CCACTGGGAAACCGACCTGACAGCCCTGGCCCAATGGCTGTTTTCCCAAGGATCAGTTCCCTGGA	1952
Db	540	CCACTGGGAAACCGACCTGACAGCCCTGGCCCAATGGCTGTTTTCCCAAGGATCAGTTCCCTGGA	599
Qy	1953	GGAAGAGGCTCTGGCCCTGCACTCCGCTGTGTCCCGACACACGTCTGACCGACGCCGC	2012
Db	600	GGAAGAGGCTCTGGCCCTGCACTCCGCTGTGTCCCGACACACGTCTGACCGACGCCGC	659
Qy	2013	CGCCCTGTAGTTCTTGGCTGGGTCTGGAAGTGTCTGTGAGACACCTTGCCTCACCCACAG	2072
Db	660	CGCCCTGTAGTTCTTGGCTGGGTCTGGAAGTGTCTGTGAGACACCTTGCCTCACCCACAG	719
Qy	2073	GAGCGTAGGCACTTCGACAGTCCACGCTGAACATGGGAAACAAACCTGAAAGACGAGCAG	2132
Db	720	GAGCGTAGGCACTTCGACAGTCCACGCTGAACATGGGAAACAAACCTGAAAGACGAGCAG	779
Qy	2133	GCCT-CCCGGTCAAGGAGCCTCTGCTGTGCTGTGCTTCCATGACACCTCCTCTGTCTGA	2191
Db	780	GCCTCCCCGGTCAAGGAGCCTCTGCTGTGCTGTGCTTCCATGACACCTCCTCTGTCTGA	839
Qy	2192	AATATTACTGCTTGAATCTGAGACAGA-TTGGGGGTTTATTAATACTGCTTTTATCTGAG	2250
Db	840	AA-ATTACTCTTGAAATCTGAGACAGATTTGGGTTTATTAATAACTGCTTTTATCTGAG	898
Qy	2251	AACAAACGGGTTTGGA 2268	
Db	899	AACAAACGGGTTTGGA 916	
RESULT 12			
LOCUS	BU150592	935 bp	mRNA linear EST 03-SEP-2002
DEFINITION	AGENCOURT_8728400 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6339117		
ACCESSION	BU150592		
VERSION	BU150592.1	GI:22664124	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 935)		
TITLE	NIH-MGC http://imgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@ds-femail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LMC2533 row: 9 column: 22 High quality sequence stop: 675. Location/Qualifiers 1..935		
FEATURES			
Source	/organism="Homo sapiens" /mol type="mRNA"		

```

/db_xref="taxon:9606"
/clone="IMAGE:6339117"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_47"
/note="Organ: Brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

```

## ORIGIN

```

Query Match      26.7%; Score 858; DB 13; Length 935;
Best Local Similarity 98.5%; Pred. No. 3.3e-212;
Matches 908; Conservative 0; Mismatches 10; Indels 4; Gaps 4;

QY 38 GGAGCTGACATGAGACATATTACTGCTGTAACGTCGCGAGGAGGCTCTTCTGA 97
| 1 GGAGCTGACATGAGACATATTACTGCTGTAACGTCGCGAGGAGGCTCTTCTGA 60
| 98 AGAGCTGAAAAGCTGCTTACCGAGGCTCTGTATCTCTACATCCAGACAGACAGA 157
| 61 AGAGCTGAAAAGCTGCTTACCGAGGCTCTGTATCTCTACATCCAGACAGACAGA 120
| 158 CCCAGAGCTCAAGTCACAGCGGAGGAGGCTTTAAGCTTTTACAGGCTTATGAAGT 217
| 121 CCCAGAGCTCAAGTCACAGCGGAGGAGGCTTTAAGCTTTTACAGGCTTATGAAGT 180
| 218 GCTTAAAGTACCCCAACAGGAGGCTCTGTATGATATATGAGAGAGAGCTGGAAT 277
| 181 GCTTAAAGTACCCCAACAGGAGGCTCTGTATGATATATGAGAGAGAGCTGGAAT 240
| 278 GGAAGATGAGAGGCTTGTGAAAGAGAGAGACCCCTGCTGAATTCGAGAGAGTTTGA 337
| 241 GGAAGATGAGAGGCTTGTGAAAGAGAGAGACCCCTGCTGAATTCGAGAGAGTTTGA 300
| 338 GCGGCTGACAGAGAGAGAGAGAGAGAGATGTCAGAGCGGACCAATCCCAAGGAAAC 397
| 301 GCGGCTGACAGAGAGAGAGAGAGAGAGATGTCAGAGCGGACCAATCCCAAGGAAAC 360
| 398 GATCAGCGTTGAGATGATCCACCGACCTTTTGAATCGCTATGATGAGAGATGAAGA 457
| 361 GATCAGCGTTGAGATGATCCACCGACCTTTTGAATCGCTATGATGAGAGATGAAGA 420
| 458 TGTGTCCGCGAGTATGCTTCCGAGATTTAAATTAATGCAATATCCAGTCCAT 517
| 421 TGTGTCCGCGAGTATGCTTCCGAGATTTAAATTAATGCAATATCCAGTCCAT 480
| 518 TGAGGACCTTTGACAGCGACAGACAGGACCTCTCGAAGGCTCTCAACCGAGAA 577
| 481 TGAGGACCTTTGACAGCGACAGACAGGACCTCTCTCGAAGGCTCTCAACCGAGAA 540
| 578 TGAGGATGAGAGAGGTTTCACTTAATCTTGCCTCAGACAGATTAATTCGCAAGGAGAT 637
| 541 TGAGGATGAGAGAGGTTTCACTTAATCTTGCCTCAGACAGATTAATTCGCAAGGAGAT 600
| 638 GGAAGAGTTGGAATTTGAGAGTGAAGACTCAAGGAGGCTTTGTTGCTCAAGCTGT 697
| 601 GGAAGAGTTGGAATTTGAGAGTGAAGACTCAAGGAGGCTTTGTTGCTCAAGCTGT 660
| 698 CGGTAAATCTACACCAAGATGCTTTTGAGACAAACATGCTCTGCAAGTTTCATCCG 757
| 661 CGGTAAATCTACACCAAGATGCTTTTGAGACAAACATGCTCTGCAAGTTTCATCCG 720
| 758 TGGAATCCGACCCGCGCTGACACTGTCTAGCTGGAACCTTGAAGAGAGACCGTGG 817
| 721 TGGAATCCGACCCGCGCTGACACTGTCTAGCTGGAACCTTGAAGAGAGACCGTGG 780
| 818 -CTACTGCAAGTGGCGATGGGATATCCAGTCAAGCATGAACACTGACATCGT-CCGAGAC 875

```

```

| 781 CCTACTCGACATGGGAGATGGGATATCAGTCAAGCATGAACACTGACATCGTCCGAGAC 840
| 876 ACTAAACAGACACTTCACTGT-GGCCCTGAGCTGGGAATCCGCACTCC-TTTGAC 933
| 841 ACTAAAGAGACACTTCACTGTGGGCGCTGAGCTGGGAATCTCACTCTTTGAC 900
| 934 TGATCAGCTATCAGCAAAATT 955
| 901 TGATCAGCTATCCGACAAAT 922

```

## RESULT 13

B0062383

LOCUS

DEFINITION

AGENCOURT 6827086 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5923853

5', mRNA sequence.

ACCESSION

B0062383

VERSION

B0062383.1 GI:19889163

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

NIH-MGC http://mhc.nci.nih.gov/.

1 (bases 1 to 1017)

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Lou Staudt

CDNA Library Preparation: Rubin Laboratory

DNA sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHCW2093 row: a column: 06

High quality sequence start: 37

High quality sequence stop: 580.

Location/Qualifiers

1..1017

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5923853"

/tissue\_type="lymphoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_99"

/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGAG(G). Size-selected &gt;500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

library."

## ORIGIN

```

Query Match      26.7%; Score 857.2; DB 13; Length 1017;
Best Local Similarity 95.5%; Pred. No. 5.5e-212;
Matches 970; Conservative 0; Mismatches 33; Indels 13; Gaps 8;

QY 5 GGTTCGGAAGATGGGAGGCGCTTTGAGGAGAGAGAGCTGACATTTGAAGATATTATCTC 64
| 1 GGTTCGGAAGATGGGAGGCGCTTTGAGGAGAGAGAGCTGACATTTGAAGATATTATCTC 57
| 65 GTTGTGAAAGCTGCGGAGGAGGCTCTTCTGAAAGCTGAAAGCTGCTACCGAGGCT 124
| 58 GTTGTGAAAGCTGCGGAGGAGGCTCTTCTGAAAGCTGAAAGCTGCTACCGAGGCT 117
| 125 CTGTATGCTTCACTCAGACAGACAGAGACCCAGAGCTCAAGTCAAGCGGAAAG 184

```

```

Db      118 CTGATGCTACCATCCAGCAAGACAGAGACCCAGAGCTCAAGTACAGGGGGAAACG 177
Qy      185 ACTGTTTAACCTTGTTCACAGGCTTATGAAGTCTTAGTACCCCAACCAAGGCCAT 244
Db      178 ACTGTTTAACCTTGTTCACAGGCTTATGAAGTCTTAGTACCCCAACCAAGGCCAT 237
Qy      245 CTATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
Db      238 CTATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 297
Qy      305 GAGAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
Db      298 GAGAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357
Qy      365 GAGATTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
Db      358 GAGATTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417
Qy      425 CCTTTTGTATGCTATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAGATGATGAG 484
Db      418 CTTTGTGATGCTATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAGATGATGAG 477
Qy      485 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 544
Db      478 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 537
Qy      545 AGCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
Db      538 AGCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
Qy      605 TGCGCTGACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 664
Db      598 TGCGCTGACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 657
Qy      665 CCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
Db      658 CCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
Qy      725 GACAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 784
Db      718 GACAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 777
Qy      785 CCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842
Db      778 CCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
Qy      843 CAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db      838 CAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
Qy      901 --CCCTGAGCTGGAATTCCTCACTCTCTTGTGACT-GATCAGCTATCAGCAAAATTC 957
Db      898 GCGCTGAGCTGGAATTCCTCACTCTCTTGTGACTGATCAGCTATCAGCAAAATTC 957
Qy      958 AAGATGAC---GATCAGCTGCTGTAAGAGATCCCTCAAGAGAGAGAGAGAGAGAGAG 1010
Db      958 AAGATGACCGATTCGACTCCGCTGGAAGAGATCCCTCAAGAGAGAGAGAGAGAGAGAG 1013

```

```

RESULT 14
LOCUS   B0063825 1018 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT 6850301 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5925479
5', mRNA sequence.
ACCESSION B0063825
VERSION   B0063825.1 GI:19891913
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1018)

```

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: sgaps-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: LCM2097 row: d column: 24
High quality sequence stop: 758.
FEATURES
Source
Location/Qualifiers
1..1018
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5925479"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
ORIGIN
Query Match 26.7%; Score 856.6; DB 13; Length 1018;
Best Local Similarity 96.8%; Pred. No. 7.9e-212;
Matches 928; Conservative 0; Mismatches 21; Indels 10; Gaps 5;
Qy      7 TTGGGAAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66
Db      1 TTGGGAAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy      67 TGCTGAACGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
Db      61 TGCTGAACGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy      127 GTATGCTCTACATCCAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 186
Db      121 GTATGCTCTACATCCAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 180
Qy      187 TGTTTAACTTGTTCACAGAGCTTATGAGTCTTATGACCCCAACCAAGGCCATCT 246
Db      181 TGTTTAACTTGTTCACAGAGCTTATGAGTCTTATGACCCCAACCAAGGCCATCT 240
Qy      247 ATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306
Db      241 ATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      307 GAAACCCCTGCTGAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
Db      301 GAAACCCCTGCTGAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy      367 GATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Db      361 GATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      427 TTTTGTATGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
Db      421 TTTTGTATGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      487 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 546
Db      481 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540

```



QY 547 CCATCTCTCTGGAAGCCTCTCAACCAGATGGAATGAGGAGGTTCCATTAACTTTG 606  
 541 CCATCTCTCTGGAAGCCTCTCAACCAGATGGAATGAGGAGGTTCCATTAACTTTG 600  
 QY 607 CGCTCAGACGAGTAACTTCCGCAAGAGATGGGAGAGTTGAAATTTGAGCTGAGACC 666  
 601 CGCTCAGACGAGTAACTTCCGCAAGAGATGGGAGAGTTGAAATTTGAGCTGAGACC 660  
 QY 667 TAAGAGGAGCCTTTGTTCCGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA 726  
 661 TAAGAGGAGCCTTTGTTCCGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA 720  
 QY 727 CAACAACTGTGCTGTCAGATTTTCAATCCGTTGAATCCGACCCGGCTTACCACTGTCC 786  
 721 CAACAACTGTGCTGTCAGATTTTCAATCCGTTGAATCCGACCCGGCTTACCACTGTCC 780  
 QY 787 TAGCTCGGAACCTAGACAGAGACACCGT-GGGCTACCTGCAATGGCGAT-GGGGTAATCCA 844  
 781 TAGCTCGGAACCTAGACAGAGACACCGTGGGGGCTACCTGCAATGGCGATGGGGTAATCCA 840  
 QY 845 GTAGAGCAGTAACTAGATGCTGCTCGAGACACTAAAC--AGCCACTTCACTGTGGCC 902  
 841 GTAGAGCAGTAACTAGATGCTGCTCGAGACACTAAACCCAGCCACTTCACTGTGGCC 900  
 QY 903 C-TGCAGCTGGAA-----TCCCTCACTCTTTCAGCTGATCAGCTATCAGACAAATT 955  
 901 CGTGACGCTGGAAATCCCTCACTCTCTTTTGACATCGAATCACTATCAGACCAAAAT 959

RESULT 15  
 AL575646/c 989 bp mRNA linear EST 01-JUN-2003  
 LOCUS AL575646 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 DEFINITION clone CS0D10691P21 3-PRIME, mRNA sequence.  
 ACCESSION AL575646 GI:31313953  
 VERSION AL575646  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 989)  
 Li,W.B., Gruber,C., Jesse,J., and Polayes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 16, 2001 this sequence version replaced gi:12937012.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 271.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1069CH1INP1&cluster=271.r. Contact :  
 Feng Liang Email: fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0D1069CH1INP1.

## FEATURES

source

1. 989  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D10691P21"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 26.5%; Score 852.8; DB 9; Length 989;  
 Best local Similarity 95.2%; Pred. No. 7.7e-211;  
 Matches 902; Conservative 13; Mismatches 28; Indels 4; Gaps 4;

QY 2238 CTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCGTTCTTTTCCCACTCCAGAG 2297  
 945 CTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCGTTCTTTTCCCACTCCAGAG 887  
 Db 2298 CTGCTCAAGTCAATTCACCGGCGCCCTCGGCTTTGGGACAGGGGTAGTAACTCCGATCC 2357  
 886 CTGCTCAAGTCAATTCACCGGCGCCCTCGGCTTTGGGACAGGGGTAGTAACTCCGATCC 827  
 QY 2358 CAGGGCTTACCTGACACAGGTGCTTCCGTTATCCCGTTGGGAAAACGCCCTGCAAC 2417  
 826 CA-GGCTTACCTGACACAGGTGCTTCCGTTATCCCGTTGGGAAAACGCCCTGCAAC 768  
 QY 2418 AGCGGCTTACCTGACCTGCTGCTGCTTCCCTCAGCCGCTGACCAACCACTTCAGAGTCAG 2477  
 767 AGCGGCTTACCTGACCTGCTGCTGCTTCCCTCAGCCGCTGACCAACCACTTCAGAGTCAG 708  
 Db 2478 TGCTGGGCAAGGGGAGCTTCAAGAGACAGGACAGGCGCTTGGCAAGATGACACAC 2537  
 707 TGCTGGGCAAGGGGAGCTTCAAGAGACAGGACAGGCGCTTGGCAAGATGACACAC 648  
 QY 2538 CCAACCCAAAGGCGTGAACCCCGAGCCCGGCTGTTAGTACCAAGAGGTGACACTGCAAC 2597  
 647 CCAACCCAAAGGCGTGAACCCCGAGCCCGGCTGTTAGTACCAAGAGGTGACACTGCAAC 588  
 Db 2598 TCCTCGCTCTGCAAGTCAAGCTGCTTCAAGAAACCAAGGCTGTGCTCCGAGCTT 2657  
 587 TCCTCGCTCTGCAAGTCAAGCTGCTTCAAGAAACCAAGGCTGTGCTCCGAGCTT 528  
 QY 2658 TCCTTCAGACCTTCTCTTCCAGCTGACGTCACCTTGGAGTGCAGAAATGACAGGAGCTTGAAGC 2717  
 527 TCCTTCAGACCTTCTCTTCCAGCTGACGTCACCTTGGAGTGCAGAAATGACAGGAGCTTGAAGC 468  
 Db 2718 CCTCCAGCGCTGAGACCTCGCTGAGTAAAGTTACGTAGAGCTGTCTCTCGGGGCTT 2777  
 467 CCTCCAGCGCTGAGACCTCGCTGAGTAAAGTTACGTAGAGCTGTCTCTCGGGGCTT 408  
 QY 2778 GGAAGTGGCAGCCATCAGTGTCTCTTCTGACACCTTCGAGACAGCGCCGACAGGTGT 2837  
 407 GGAAGTGGCAGCCATCAGTGTCTCTTCTGACACCTTCGAGACAGCGCCGACAGGTGT 348  
 Db 2838 GGCTGAGACAGTGGGCGGGGGGCGCCCAAGCTGCGCGGCTTCAGGCCACCAACAGCT 2897  
 347 GGCTGAGACAGTGGGCGGGGGGCGCCCAAG-TGGCGGCGCTTCAGGCCACCAACAGCT 289  
 QY 2898 GTTGTGAGAGTCAAGGCAACCTCCCGAGCACTGTATCTGAGTAAACGCTTAAGAACCTCC 2957  
 288 GTTGTGAGAGTCAAGGCTTCCCTCCCGAGCACTGTATCTGAGTAAACGCTTAAGAACCTCC 229  
 QY 2958 TTCTCTGTGTTTGAAGACAGTTGGGTTGTCAATTCTGTAACATTTCATCTCAATTTT 3017  
 228 TTCTCTGTGTTTGAAGACAGTTGGGTTGTCAATTCTGTAACATTTCATCTCAATTTT 169  
 Db 3018 TTAAGAAAGGTTTCTGACGCGCCCAAGCGCCGAGCGCGGTTGAGCGTGGTTCAGT 3077  
 168 TT-TAAGAAAGTTCCTGACGCGCCCAAGCGCGGAGCGCGGTTGAGCGTGGTTCAGT 110  
 QY 3078 AGCTTGGGCGCGGAGCTTCCCTGCGCTCTGACCGCAGGTGCTTGTGGGACCCATCTTC 3137  
 109 AGCTTGGGCGCGGAGCTTCCCTGCGCTCTGACCGCAGGTGCTTGTGGGACCCATCTTC 50  
 Db 3138 TCGCTTCAATTTGCACTGACCTGTAACAGAGCACTCAACCAATTA 3184  
 49 TCGCTTCAATTTGCACTGACCTGTAACAGAGCABTNVMMCBAAATTA 3

Search completed: July 13, 2004, 14:50:21  
 Job time : 7868 secs